

---

Electronic Thesis and Dissertation Repository

---

8-7-2019 1:00 PM

## Methodological Challenges for Empirical Approaches to Ethics

Christopher Shirreff  
*The University of Western Ontario*

Supervisor  
Brennan, Samantha J.  
*University of Guelph*

Graduate Program in Philosophy  
A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of  
Philosophy  
© Christopher Shirreff 2019

Follow this and additional works at: <https://ir.lib.uwo.ca/etd>



Part of the [Philosophy Commons](#)

---

### Recommended Citation

Shirreff, Christopher, "Methodological Challenges for Empirical Approaches to Ethics" (2019). *Electronic Thesis and Dissertation Repository*. 6341.  
<https://ir.lib.uwo.ca/etd/6341>

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact [wlsadmin@uwo.ca](mailto:wlsadmin@uwo.ca).

## Abstract

The central question for this dissertation is, how do we do moral philosophy well from within a broadly naturalist framework? Its main goal is to lay the groundwork for a methodological approach to moral philosophy that integrates traditional, intuition-driven approaches to ethics with empirical approaches that employ empirical data from biology and cognitive science. Specifically, it explores what restrictions are placed on our moral theorizing by findings in evolutionary biology, psychology, neuroscience, and other fields, and how we can integrate this data while still offering a fully normative account of ethics. To that end, the dissertation explores the methodological assumptions behind both traditional and more empirical or experimental approaches to ethics, to find where these assumptions cannot be properly supported and need to be re-examined. Chapter 1 explores the science-based objections that have been raised to traditional approaches, with a particular focus on questions concerning the reliability of moral intuitions. Chapter 2 examines three fundamental assumptions supporting the empirical approach to moral philosophy, and how those assumptions ultimately do not fit with how we ought to understand the project of ethics. In chapter 3, I discuss the fact/value distinction, and the restrictions that are placed on our moral theorizing by our commitment to this distinction. Chapter 4 offers a defense of ‘companions in guilt’ arguments, and uses these arguments to draw an analogy between moral philosophy and epistemology that will be used to help defend moral philosophy against empirical debunking arguments. Chapter 5 explores the ways in which epistemologists’ methods of incorporating empirical data into their research while maintaining the normativity of their accounts can be adapted to allow moral philosophers to do the same, and brings together the various methodological concerns addressed up to this point to lay out a methodological approach I call empirically-informed moral philosophy.

## Keywords

Metaethics, Normative Ethics, Moral Philosophy, Moral Intuition, Experimental Philosophy, Fact/Value Distinction.

## Summary for Lay Audience

Moral philosophers have long been resistant to using scientific data in their philosophical work, owing to a commitment to the idea that descriptions of the world (claims that tell us what *is* the case) cannot tell us anything about what *ought* to be the case, or what we *ought* to do, which is what moral philosophers are concerned with. Recently, however, some philosophers have taken note of research from psychologists, neuroscientists, biologists, and others that seem to have implications for our understanding of ethics. Some of these philosophers have even begun to conduct their own experiments to learn more about how people actually think about ethics and what is going on in the brain when people think about moral questions. This has led to two approaches to ethics: what I call the “traditional approach,” which continues to rely on pure intuition and reasoning, and what I call the “empirical approach,” which relies heavily on scientific data. In this thesis, I look at these two approaches, and ask what each of them must think about how we should understand ethics to justify their approach. I argue that neither approach can give a fully satisfactory account of ethics, as both rely on background assumptions that cannot be entirely justified. I then consider general restrictions on how we approach moral philosophy, and how this can help to deal with these issues. I look at how other areas of philosophy have contended with similar challenges, and ways in which their solutions can be used to address the problematic background assumptions of both the traditional and empirical approaches. Applying these lessons, I present my approach, striking a balance between the two existing approaches, which I call the “empirically-informed” approach. This approach makes use of scientific data, but does not rely on it as heavily as the purely empirical approach, and so does not face the same issues as either other method.

## Acknowledgments

This project is the result of a long and complex process, and a full listing of all those who have directly or indirectly aided in its completion is far too extensive to list here. However, I must give thanks to some people who were absolutely essential to its completion, and offer apologies to any I fail to thank personally.

I must begin by thanking my supervisor, Samantha Brennan, for her support advice, often unfathomably fast feedback, and, above all, patience. This thesis simply would not exist without her guidance, and for that, I am endlessly grateful. I also want to thank Tracy Isaacs for her feedback on earlier versions of these chapters, which considerably improved the quality of this work.

Thanks also to my examination committee, Tracy Isaacs, Robert Corless, Jacquie Burkell, and Carla Fehr. Their questions and comments have both helped me to clarify ideas in this thesis, and provided future avenues of inquiry for this project.

Further thanks go to the entire philosophy community at Western and the Rotman Institute for providing discussion, support, and a place for these ideas to develop, with special thanks to Gillian Barker and Anthony Skelton.

I would like to thank my friends and colleagues, Dylan Gault and Emerson Doyle, who were the first readers of every chapter of this project, and whose detailed comments and support were invaluable. I am also grateful to Jon Life, Danny Booth, Shari Monner, Patrick Clipsham, Rebecca MacIntosh, Katy Fulfer, Jessey Wright, Helana Hope, Robert Foley, Dan Hicks, Kevin Godbout, Torin Doppelt, Martin Lefebvre, and Madison Bettel for their friendship, encouragement, support, and advice, both academic and otherwise. Particular thanks go to Fred Banville for both his friendship and antagonism.

I had the great privilege of working with the Society of Graduate Students during my time at Western, and offer my thanks to all the friends and colleagues I met and worked with during my time there.

I must give special thanks to the late Jackie Brunning, whose impact on my career was inestimable. Without her influence at UTM, I may never have pursued philosophy in the first place. I owe a tremendous debt to her for her guidance in the earliest days of my philosophical life.

I also wish to thank my family for all their support: my mother, Rhonda; my father, John; my aunt, Joanne; and, of course, my brother and fellow philosopher, Patrick, for a lifetime of discussion, debate, and argument on matters both philosophical and trivial.

Finally, it would be terribly remiss of me to not thank Freya, my semi-loyal feline companion. Although she will not at any point read, understand, or have even the slightest interest in any part of this thesis, her constant companionship (and more than occasional interference) is nevertheless greatly appreciated.

# Table of Contents

Abstract .....	i
Summary for Lay Audience .....	ii
Acknowledgments .....	iii
Table of Contents .....	v
Introduction .....	1
Chapter 1 .....	5
1 The Empirical Challenge .....	5
1.1 Intuitions and Moral Intuitions .....	8
1.2 Cognitive Bias .....	11
1.3 The Darwinian Dilemma .....	22
1.4 Thought Experiments and Moral Counterexamples .....	27
1.5 Conclusion .....	34
Chapter 2 .....	36
2 Methodological Assumptions of an Empirical Approach to Ethics .....	36
2.1 The Empirical Assumptions .....	37
2.2 Moral Expertise .....	38
2.3 Describing or Prescribing Norms? .....	50
2.3.1 The Descriptive Account .....	51
2.3.2 The Moral Project .....	56
2.4 Moral Debunking .....	61
2.5 Conclusion .....	67
Chapter 3 .....	69
3 Facts and Values .....	69
3.1 Philosophy of Science and the Fact/Value Distinction .....	70

3.2 The Is/Ought Gap.....	72
3.3 Single-Premise Arguments .....	75
3.4 Multi-Premise Arguments.....	82
3.5 Conclusion .....	89
Chapter 4.....	92
4 Companions in Guilt Arguments .....	92
4.1 Defending Companions in Guilt Arguments .....	94
4.2 Limits of Companions in Guilt Arguments .....	104
Chapter 5.....	110
5 Empirically-Informed Moral Philosophy .....	110
5.1 Companions in Defense .....	111
5.2 Moral Intuitions Revisited .....	120
5.3 Resisting Cognitive Bias.....	121
5.4 Biofatalism.....	124
5.5 Fallibilism and Thought Experiments.....	127
5.6 The Methods of Ethics .....	132
5.7 Conclusion .....	137
Bibliography .....	140
Curriculum Vitae .....	148

## Introduction

It is an interesting time to be doing work in ethics. The theory of evolution by natural selection is as well-supported as any other idea in contemporary science, psychology and neuroscience have given us unprecedented access to the structure and functioning of the human mind, and advances in neuroimaging have allowed us to see exactly what parts of our brain are active as we puzzle over a moral dilemma. Unsurprisingly, philosophers have begun to take note of these discoveries and incorporate them into their research. But what does this mean for those of us working in moral philosophy? One of the first things any young ethicist is taught is that there is a fact/value distinction, and that moving from a descriptive claim to a normative one is a fallacy. Nevertheless, in recent years philosophers, including Jesse Prinz (2007a, 2007b, 2008), Patricia Churchland (2011), and Joshua Greene (2008), among many others, have put forward accounts of ethics that make substantial use of empirical data to justify their claims, often including criticisms of non-empirically minded ethicists for failing to attend to relevant scientific data.

This dissertation examines the growing debate between empirical and traditional approaches to moral philosophy and the background assumptions made by those on each side of the debate, with a focus on the understanding of norms they are employing, and how they can balance normativity and empirical adequacy. The strongly empirical approach fails to offer any properly normative content, instead settling for mere descriptions of people's preferences and what leads to adopting those preferences. As such, it cannot answer any of the questions we pose about what we should do or prefer, which form the fundamental questions of ethics. On the other hand, the non-empirical approach leads to theories at odds with established scientific facts, and thus incompatible with the naturalist position that we should believe what science tells us about the world. If these are the only two options, then the conclusion seems to be that we simply cannot do ethics, but this is a profoundly unsatisfying outcome. We are not going to stop asking questions about how we ought to behave or relate to each other, and these seem to be meaningful questions, yet our current approaches to ethics cannot make sense of this from within the framework of our currently-accepted sciences. The question before us



then is this: how do we do moral philosophy well from within a broadly naturalist worldview, making use of our best-defended scientific theories? To answer this question, I attempt to navigate a space between the two positions to articulate an understanding of norms which can move beyond mere descriptions of our preferences, but still fit with our broadly naturalist understanding of the world.

Following Patrick Clipsham (2014), I take the debate between proponents of the traditional and empirical approaches to be a methodological one. That is to say, the debate is not about what the facts of the matter are. The empirical ethicist can accept that the traditionalist's intuitions are exactly as she says they are, and the traditionalist does not need to argue that any of the studies cited by the empiricist are somehow flawed. They do not argue that these pieces of evidence are mistaken; they argue that they are not relevant pieces of evidence regardless of whether they are right or wrong. What is at issue is what counts as evidence in the first place, and what weight should be given to various facts. The traditional approach, often citing the fact/value distinction, heavily discounts the evidential value of the empirical studies cited by the empiricists. The empirical approach, on the other hand, discounts the value of the philosopher's moral intuitions, usually arguing that they are somehow out of touch, or otherwise inaccurate.<sup>1</sup> So, we can safely treat this as a debate not about what the facts are, but which facts are or are not relevant, and what weight we give to various facts in our deliberation: a methodological debate.

Since this debate is methodological, concerned with what counts as evidence for claims in moral philosophy and the aims of moral philosophy, I am interested mainly in the methodological assumptions that inform each approach, and what these assumptions reveal about the major views regarding the nature of norms. On the empirical side of the debate, I examine how philosophers have attempted to sidestep or argue against the fact/value distinction to support the incorporation of scientific data into a normative project, and what assumptions are being made about the nature of moral inquiry and what

---

<sup>1</sup> But, importantly, they do *not* argue that the philosopher is mistaken in her evaluation of what her intuitions are in the first place.

might count as evidence in favour of moral claims. On the traditional side, I explore how philosophers have addressed challenges posed by empirically-minded philosophers (or philosophically-minded scientists) to traditional approaches to ethics. I look particularly at how these accounts have attempted to accommodate challenges from scientific literature that seem to undermine the reliability of the intuitions which can often serve as bedrock for moral theorizing. The points raised by Tversky and Kahneman (1981) about the framing effect, for example, show that what we take to be the morally correct act depends more on how the question is presented than any of what we typically take to be morally relevant facts. How ethicists respond when scientific data seem to conflict with their moral theories, and what this tells us about background assumptions regarding the relationship between ethics and science, will be of particular interest.

This dissertation has five chapters. In the first two, I explore the methods of the traditional and empirical approaches to philosophy and the objections that have been raised against these positions. I focus on what these approaches reveal about the background assumptions that inform these approaches. I argue that both the traditional and empirical approaches rely on problematic methodological assumptions which undermine the efficacy of the approaches. In the third chapter, I look to the methodological commitment that is at the heart of any debate between empirical and traditional ethicists: the fact/value distinction. I argue that despite numerous attempts to provide is/ought bridging arguments over many years, there is still no compelling argument that should convince us that it is possible to derive normative conclusions from purely descriptive premises. In the fourth and fifth chapters, I present the methodological restrictions that our investigations up to that point have created for us, and how we can deal with this challenge. To that end, I defend companions in guilt arguments that attempt to defend moral normativity by pointing to the analogy to epistemic normativity. With the relationship between epistemology and ethics established, I look to ways in which epistemologists have addressed the same sort of empirical challenges that ethics faces, and argue for ways that moral philosophers can make use of similar methodological approaches. Finally, I conclude that, although the challenge from empirical data presents significant challenges for moral philosophy, there are methodological tools available that

allow us to offer prescriptive accounts of morality while keeping our theories in line with the best available scientific data.

Before beginning, however, there are two aspects of my own methodological approach that are worth mentioning. First, I mentioned above that I am concerned with looking at moral philosophy from within a naturalist worldview, but it is worth clarifying what I meant by that. Prinz has offered an outline of four kinds of naturalism (Prinz 2007a, 2-3), but for the purposes of this argument, I do not need to commit to quite as much as Prinz does. All I mean by a naturalist worldview is the attitude that the world is more or less as described or modeled by our best-established scientific theories,<sup>2</sup> and that we should avoid appealing to supernatural explanations.<sup>3</sup> Second, I am interested in how philosophers have analyzed and used empirical data to support moral arguments, not necessarily in how they collect their data in the first place. Since – as I mentioned above – this debate is ultimately not really about facts, but about what counts as evidence and how to weigh that evidence, I am generally not interested in critiquing the design of various experiments, or questioning the validity of the raw experimental data. I am happy to generally assume that any given experiment was designed correctly, and that the data gathered are accurate. Instead, what is of interest is how the data have been analyzed and interpreted, what conclusions have been drawn from them, and whether or not those conclusions are well-supported.

---

<sup>2</sup> Although this assumption will periodically be bracketed for the sake of argument.

<sup>3</sup> e.g., explaining some phenomenon as being literally magic, a miracle, direct divine intervention, a result of strange, mystical forces, or other similar explanations.

## Chapter 1

### 1 The Empirical Challenge

In the broadest terms, there are two separate but closely-related objections raised against traditional approaches to ethics and the role of intuition in such approaches.<sup>4</sup> According to the first objection, current approaches to philosophy do not sufficiently attend to questions about what kind of creatures we are, and how these facts inform the origins of our intuitions (Knobe & Nichols 2008, 3). According to the second objection, because of the isolated nature of armchair philosophizing, we cannot know with any degree of confidence that our intuitions about particular cases are reliable without understanding the psychological or cultural mechanisms underlying those intuitions (*ibid.*, 7-8). The empirical challenge, then, is not that moral philosophy as it is typically practiced is ineffective, or has “stalled” and is incapable of making progress. Rather, the challenge is that the methods employed by moral philosophers do not pay enough attention to the empirical realities of the kind of evolved creature that humans are, and what this means for the reliability of any philosophical theory that relies heavily on the use of intuitions.<sup>5</sup>

It is not difficult to see at least some of the force of this challenge. If we are relying on intuitions as evidence or grounding for our theories, we want those intuitions to be robust and reliable. If not, their value as evidence declines sharply. To some extent, testing that our intuitions about some particular case are not just our own idiosyncratic views is a common behaviour for nearly all philosophers. Every time we ask a colleague what they think about a thought experiment we are considering for a paper, or whether or not a claim being made needs further defense, we are testing whether or not our intuitions are shared by others, or if we need to reconsider our own intuitions. The empirical

---

<sup>4</sup> Some of these objections apply well beyond moral philosophy, but for the purposes of this chapter, how these apply to other fields can be set aside.

<sup>5</sup> This is not to suggest that intuitionism is all there is to it. Robert Noggle (2001), for instance, has proposed an account of morality based not just in intuitions, but in theories about human nature, which he calls the “nature to morality methodology” (532). However, discussion of such approaches is outside the scope of this chapter, and will be discussed further in chapter 4.

approach to moral philosophy suggests expanding this, testing our intuitions not just against those of colleagues and friends, but against the wider population.

In this chapter, I discuss the methodological approaches employed by traditional approaches to ethics, and how moral philosophers are able to respond to the challenges put forward by proponents of the empirical approach. I focus on the role of intuitions in moral philosophy. This chapter has four main sections. In section one, I argue for the indispensability of intuitions to traditional accounts of moral philosophy, and the need for those intuitions to be reliable if the traditional approach is to be successful. In the second section, I present the challenges raised for the reliability of our moral intuitions by cognitive biases, focusing on the framing effect. I argue that psychological research on the framing effect has shown that even some of our more fundamental moral intuitions can be distorted by non-moral factors. This distortion, combined with our current inability to know the exact scope of the problem, serves to undermine our confidence in moral intuitions. In section three I look at Sharon Street's "Darwinian Dilemma" (2006) and the epistemic difficulties of attempting to preserve moral intuitions and realism in the face of scientific claims. I argue that this provides us with another reason to be skeptical of our moral intuitions that cannot be resisted without descending to an unacceptable skepticism about science. This section also looks at what limits exist on the abilities of traditional moral philosophers to respond to the criticisms and challenges of their empirical counterparts. In the fourth and final section, I look at the use of thought experiments in moral philosophy, with particular focus on their use of intuitions. I argue that, although in general thought experiments are a useful investigative tool for moral philosophers, we must be cautious of some methodologically dubious uses of them. I conclude the chapter by arguing that the challenges raised by empirical approaches to moral philosophy show that the background assumption underlying the traditional investigative tools of moral philosophy – that intuitions are a powerful and reliable investigative tool – cannot be adequately maintained, and our intuitions need to be approached with a great deal of caution.

Before beginning in earnest, there is an issue that must be addressed. It will no doubt seem immediately odd to many moral philosophers that I am discussing the

methodological tools of moral philosophy without focusing heavily (or indeed much at all) on the method of reflective equilibrium,<sup>6</sup> arguably the method most closely and uniquely identified with the study of moral philosophy. As I stated at the outset, however, the aim of this project is to look at the methodological questions on which traditional and empirical philosophers differ, especially with regard to what sorts of things count as evidence. On questions pertaining to the proper role and evidential value of philosophers' intuitions, there is a large gap between the two camps. It is far less clear, however, that there is substantive disagreement about reflective equilibrium.

One arrives at reflective equilibrium by balancing normative theories against intuitions in particular cases, making adjustments as needed, and seeking coherence between general theories and particular judgments. Although there are debates about exactly how this balance is to be achieved, and some have been critical of its apparently coherentist understanding of justification,<sup>7</sup> at its core, reflective equilibrium is merely a means of weighing evidence and adjusting theories. We weigh the evidence (in this case, intuitions) against our existing theories to arrive at an understanding of what our theory says about this case, and potentially what our theory gets wrong answer in this case. Regardless of the particulars of our theories of justification, we should all agree that weighing our theories against the evidence available to us, and making revisions based on how they balance is – broadly speaking – how we ought to approach things, whether in moral philosophy, science, or anywhere else. The disagreement between empirical and traditional philosophers can best be understood not as a disagreement about whether or not reflective equilibrium is a useful methodological tool, but rather one over what weight different pieces of evidence ought to be given in one's attempts to reach reflective equilibrium. The empirical ethicist believes that more weight should be given to empirical data supported by our best science, while a more traditional approach will tend to discount such descriptive claims fairly heavily. While this disagreement involves, and could be said to revolve around, reflective equilibrium, it would be a mistake to take it as

---

<sup>6</sup> Cf. Rawls 1971.

<sup>7</sup> For an example of both criticisms, see Arras 2007.

a debate about the value of reflective equilibrium. As this debate is over what weight to give various pieces of evidence, and not over the tool used to analyze and balance that evidence, we may safely set aside any concerns about the method of reflective equilibrium from our present discussion.

## 1.1 Intuitions and Moral Intuitions

Intuitions are a major tool of philosophical inquiry, serving as support for theories or as their foundation. And while they are not the be all and end all of conceptual analysis, they are certainly a central tool. When we ask if a certain term or concept is appropriately used in a given context, we are checking what our intuitions – as competent users of the language – tell us about the particular case. For instance, if I say that the concept “stool” properly applies to a barstool, but not to an office chair, claiming that calling one a stool would seem odd to most English speakers – as opposed to citing a list of necessary and sufficient conditions for something to count as a stool and pointing out that the qualities of office chairs don’t match the list – I am citing an intuition about the use of language and the concept in question. Similarly, when we provide a counterexample to an analysis of a concept in terms of some set of necessary and sufficient conditions by pointing to a case that satisfies those conditions but fails to fall under that concept, or vice versa, this is also pointing to an intuition about concepts. Gettier’s (1963) famous counterexamples to the justified true belief (JTB) account of knowledge operate by citing instances in which an agent has a justified, true belief about a situation, but nevertheless our natural intuition about the case is to deny that they actually have knowledge.

It is important to note that it is not merely a coincidence or historical accident that intuitions have come to occupy such a central role in conceptual analysis. If we are attempting to analyze concepts down to their necessary and sufficient conditions, we need some set of prototypical instances of the concept, or intuitive features of the concept. If we want to know how to analyze the concept, “sandwich,”<sup>8</sup> for instance, there

---

<sup>8</sup> At time of writing, this is a live debate, often focusing on whether or not a hot dog is a sandwich (Cf. Deutsch 2015), and frequently resulting in both intuition-probing, and a taxonomy of “sandwich” drawn from that intuition-probing (Cf. Garber 2015).

are two possible approaches. The first is to consider a group of things that are uncontroversially or obviously sandwiches – which is certainly a matter of probing our intuitions, since there can be no other conditions on which to base this – then try to determine what they have in common and what differentiates them from things that are obviously not sandwiches. The other possible approach is to stipulate a set of qualities that make something a sandwich, then find things that have those qualities. There are two ways to arrive at the list of qualities. The first is to consider things that are intuitively connected to “sandwichness”. The alternative is to stipulate that “sandwich” means  $x$ ,  $y$ , and  $z$ , which has the advantage of avoiding any reliance on intuitions or presuppositions, but has the disadvantage of potentially being entirely disconnected from how the term is actually used in ordinary language.

As a further complication, it is highly unlikely that one would even be able to come up with a stipulative definition of a term without being influenced by one’s linguistic intuitions. To be a competent speaker of a language is to have some sense of how terms in the language tend to be used. In that case, absent deliberately defining it in a way that does not fit intuitions about ordinary use<sup>9</sup> – a choice that is unlikely to be useful in moving any discussion forward – we are going to be influenced by our sense of what are and are not appropriate uses of the term. Continuing with the sandwich example, I know without needing to think about it<sup>10</sup> that a clubhouse is a sandwich and a pizza is not. So, in trying to come up with a stipulative definition of “sandwich,” even if I try not to base my definition on such intuitions, this is nevertheless information to which I have access, and which will inform my thoughts. By contrast, if I am stipulating a definition for a word or concept with which I am unfamiliar, or in a language of which I am not a competent speaker, such as, “яблоко,” I can come up with a definition free of any influence from intuitions. It will, of course, be mere luck if such a definition matches up in any way with the actual use of the term, however.

---

<sup>9</sup> By defining “chair” as “a four-legged mammal with stripes”, for example.

<sup>10</sup> i.e., intuitively



If philosophy in general is reliant in intuitions, this reliance is even stronger in moral philosophy. Non-normative areas of philosophy can at least sometimes seek justification or confirmation of their intuitions by checking whether or not the world matches up to their intuitions. A philosopher of science, for example, can look to actual scientific practice to see if any tools of explanation in science seem to contradict the intuitions on which she has based her account of scientific explanation. This route is not available to moral philosophers. Since it is impossible to support a substantive moral claim from purely descriptive premises,<sup>11</sup> there are only two ways we could justify our moral claims. First, they can be justified inferentially by appealing to other claims. “Lying is wrong, therefore lying to Marcus is wrong,” or similar inferences. Alternately, they can be supported non-inferentially by intuitions. Relying solely on inferences necessarily leads to an infinite regress, since none of the moral claims in the chain can be justified by an appeal to something non-moral.<sup>12</sup> Thus, unless we accept infinite regress, somewhere down the line of justification we must cite something justified by an appeal to intuition to support our claims. The actual practice of moral philosophy proceeds from intuitions. We offer support for our theories by showing that they cohere with intuitive judgments about what is morally permissible or impermissible, and criticize others by pointing to ways in which their accounts fail to cohere with these same intuitions. One example of this would be Thomson’s use of various versions of the trolley problem and other similar hypothetical cases to show that our intuitions support the moral distinction between killing and letting die (Thomson 1976).<sup>13</sup> Similarly, certain intuitions are treated as philosophical bedrock: foundational acts of such transparent moral character that we will make sure to preserve those judgments in our theorizing. Imagine, for instance, that I were to argue utilitarianism is deeply flawed as a system of ethics because utilitarians are unable to say that torture or murder are always impermissible. I would be arguing that my

---

<sup>11</sup> A point I argue for and discuss in greater detail in chapter 3.

<sup>12</sup> For more detail on this, see Sinnott-Armstrong 2008. For a thorough argument for the impossibility of an intuition-free moral philosophy, see Sinnott-Armstrong 2006.

<sup>13</sup> Even if, as Thomson argues, this does not mean that killing always worse than letting die.

(or our) intuitions that these things are wrong are so fundamental to the concepts of right and wrong that utilitarian positions which go against these intuitions must be mistaken.

There is not necessarily a sharp line between these two uses of intuitions. Consider Elizabeth Anscombe's objection that a contractual understanding of morality will be too general and not "[descend] to such particularities as the prohibition on murder or sodomy"<sup>14</sup> (Anscombe 1958, 14). We can understand that as a claim that such prohibitions are the sort of prototypical examples of moral wrongs that serve as part of the intuitive concept of moral wrongness and thus are foundational to any reasonable moral theory. But at the same time, we can also understand it as arguing that in virtue of failing to prohibit such acts, contractualist models run afoul of our intuitions that such acts are morally unacceptable. As with the use of intuitions in philosophy in general, this is certainly not a bad thing, nor is it mere coincidence. Intuitions are unavoidable in philosophizing, moral or otherwise. The debate between traditional and empirical ethicists is not whether or not we should employ intuitions; rather, the debate is over how intuitions are employed, and how much we can rely on them.

## 1.2 Cognitive Bias

As mentioned above, most philosophers are aware of the importance of ensuring that our intuitions are reliable. We carefully consult and pick apart our intuitions, confer with colleagues, present our ideas at conferences, subject our work to peer review, in part so that we can be more certain that our intuitions are not misguided or peculiar to us.<sup>15</sup> Unfortunately, these do not necessarily cover all of the possible – or even likely – sources of error. There are a host of cognitive biases to which we are subject, and which can affect our intuitions problematic ways. Just as quick examples, according to research on confirmation bias, we are prone to seeking out information that supports what we already

---

<sup>14</sup> Setting aside the various objections to this claim, not the least of which is that there certainly is no proper moral prohibition against the latter.

<sup>15</sup> Even these are not ideal methods. As I will discuss further in the next chapter, critics of intuitions point out that consulting our philosophical colleagues means we aren't using a particularly diverse sample, which is itself a problem.

believe, and ignoring or explaining away disconfirming evidence.<sup>16</sup> Studies on the availability heuristic have shown that people tend to give more weight to newly-acquired or readily-available information than to things which are harder to recall or less salient.<sup>17</sup> Work on the fundamental attribution error points to a tendency to explain the behaviour of others based on internal factors (such as their personality or character), underestimating the role that situational factors have on their actions, while doing the opposite for our own actions.<sup>18</sup> Research on the optimism bias<sup>19</sup> has shown that people tend to underestimate the chances of something negative happening to them, which causes problems for consequentialist reasoning, as it affects our abilities to effectively perform the calculations that consequentialism demands of us.<sup>20</sup> Finally, researchers have found that people's preferences between various options can be manipulated based on seemingly-irrelevant changes to the wording of those options in what has been dubbed the framing effect.<sup>21</sup> All of these can potentially undermine our confidence in our moral intuitions, and they are just a small sample of the cognitive biases that have been recognized. This section discusses the problems these biases pose for any traditional approach, with a particular focus on the framing effect, and argues that we have good reason to doubt the reliability of our moral intuitions.<sup>22</sup>

---

<sup>16</sup> Cf. Wason 1960.

<sup>17</sup> Cf. Tversky and Kahneman 1973.

<sup>18</sup> Cf. Jones and Harris 1967.

<sup>19</sup> Cf. Sharot 2011

<sup>20</sup> I am grateful to Samantha Brennan for pointing me to this particular bias and the problems it can cause.

<sup>21</sup> For an excellent discussion on and explanation of the various cognitive biases, see Kahneman, Slovic, and Tversky 1982.

<sup>22</sup> While the complete list of cognitive biases is expansive, and continues to grow as further psychological investigations proceed, it will not be necessary to consider all possible biases for this investigation, as not all of them are morally relevant. For example, research done on implicit bias has shown various ways people can hold and act on discriminatory beliefs that they may not realize they hold, nor would they assent to if asked, including studies finding that resumes with more white sounding names being more likely to be selected for interviews than those with black, Hispanic, or Asian sounding ones (Bertrand & Mullainathan 2004, Kang et al. 2016). This affects people's judgments, but it effects implicit ones, not explicit, and our concern is with the explicit positions philosophers advance. While a philosopher may subconsciously find a CV with a typically white name more attractive than one with a typically black name, it is unlikely that this

Tversky and Kahneman (1981) show that people will make different decisions between various options depending on how those options are presented. Appropriately enough, their first case tested is also one of clear relevance to those of us working in ethics. Subjects are told that an exotic disease is going to strike that is expected to kill 600 people. They are presented with one of two pairs of research programs to combat the disease. In the first pair, they are asked to choose between program A that would save 200 people, and B that has a  $1/3$  chance of saving 600 people and a  $2/3$  chance of saving no one. In the second pair, they are asked to choose between program C that will kill 400 people, and D with a  $1/3$  chance that no one will die and a  $2/3$  chance that 600 people will die (Tversky and Kahneman 1981, 453). These two sets of options are identical in terms of their results. Programs A and C are a guarantee of 200 people living at a cost of 400 certainly dying, and programs B and D are a gamble at saving everyone with it being twice as likely that everyone will die. The only difference between them is that the first pair focuses on the number of people saved, and the second on the number who die. Despite this similarity between the cases, people give very different responses to the two sets of options. In the first set of options, people tend to prefer the sure thing, choosing to guarantee that 200 people will be saved 72% of the time, and gambling on saving everyone only 28% of the time (ibid.). When the focus is on the number of people killed, however, responses flip, with only 22% choosing the guaranteed 400 people killed, and 78% choosing to take the chance of saving everyone (ibid.).

This result is at least slightly counterintuitive. As stated above, the two sets of options should be interchangeable, and thus we should expect them to prompt more or less the same reactions from people. And yet, merely through changing how the options are presented, Tversky and Kahneman are able to prompt people to have different reactions. This is important for ethicists. The question of which research program should be pursued, and whether we should knowingly sacrifice most people to save a few or risk

---

subconscious preference would extend to her arguing that black people are less deserving of moral consideration or something similar.

saving no one if it means the possibility of saving everyone is exactly the sort of question we may debate. More importantly, the intuitions probed are exactly the sorts of intuitions a philosopher might cite to support her theory, perhaps by claiming that it is obligatory to focus more generally on research that will benefit everyone even if there is a significant risk that it will fail, because, “in this example, it is obvious that it would be wrong to condemn 400 people to death when we have the possibility of saving them.” Or similarly, arguing that our goal should always be to minimize the worst possible outcome by appealing to the “obvious” intuitive appeal of program A as opposed to B.

This causes a problem for traditional approaches to ethics that depend on the reliability of intuitions. If intuitions are crucial to our moral epistemology, we need to be concerned if they are potentially unreliable. Sinnott-Armstrong’s “*master argument*” (Sinnott-Armstrong 2008, 51 [original emphasis]) is helpful in illustrating the problem. His argument proceeds as follows:

- (1) If our moral intuitions are formed in circumstances where they are unreliable, and if we ought to know this, then our moral intuitions are not justified without inferential confirmation.
- (2) If moral intuitions are subject to framing effects, then they are not reliable in those circumstances.
- (3) Moral intuitions are subject to framing effects in many circumstances.
- (4) We ought to know (3).
- (5) Therefore, our moral intuitions in those circumstances are not justified without inferential confirmation. (ibid., 52)

By “inferential confirmation,” Sinnott-Armstrong means that the moral intuitions in question would need to be justified by being derivable from some set of premises. But as discussed above, it is impossible for all of our moral intuitions to be so derived without leading to infinite regress.

Let’s consider the *master argument*. Premise (1) is a relatively straightforward epistemic point that makes the more or less uncontroversial claim that evidence gained through an unreliable process (observations made while hallucinating being a standard

example) needs further justification from other sources if we are to be justified in believing it, and so we need say no more about it. There are several ways of defending (4) if (3) turns out to be true. Sinnott-Armstrong partially defends it by arguing that those of us who have been exposed to moral disagreement or argument will have encountered situations that could or should have caused us to realize that moral intuitions are subject to framing effects (ibid., 67). This, however, seems to overstate the case. Most of us have encountered a situation of moral disagreement in which people changed their opinions when the example or argument was presented in a different manner. But the natural response to such a situation would be to think that the changed presentation of the example made it clearer or easier to see what the correct response is. To assume that people's intuitive reactions to seemingly identical cases can be dramatically altered by, for instance, whether we talk about the number of people killed or the number saved would seem like a stretch from our ordinary encounters with moral disagreement. After all, it was relatively recently that Tversky and Kahneman first wrote about the effect. At any rate, regardless of whether or not the existence of the framing effect can be inferred from ordinary moral disagreement, there is an easier way to argue that we ought to know about it. The framing effect is not an obscure bit of psychological trivia, and much has been written about it. Further, as experts in moral philosophy, there is surely some epistemic obligation to be aware of potential defeaters or undermining factors for our beliefs, especially when those beliefs are used as justifications for other beliefs as intuitions are used in moral philosophizing. So, as a simple matter of fulfilling our epistemic obligations, we should be aware of readily-available data that should affect our confidence in the reliability of our intuitions. Therefore, if moral intuitions are subject to the framing effect, we ought to know about it. Put differently, we have an obligation to know about it *now*, when evidence is available.

This leaves us with premises (2) and (3). The truth of (2) will hinge what it means for our intuitions to be reliable and what might make them unreliable. The exact understanding is beyond the scope of this project, but at a minimum, for our moral intuitions to be reliable, they should be consistent whenever the relevant moral facts are

the same.<sup>23</sup> By definition, though, the framing effect occurs when our intuitions about a case can be changed while changing only wording. One could claim that two scenarios that are identical in every way except the words used to describe them somehow have morally different characters in light of the altered descriptions, but this is an untenable position at best, and I cannot imagine a reasonable argument to support it. Necessarily, then, the framing effect is a change to our intuitions for morally irrelevant reasons. Thus, if our moral intuitions are vulnerable to the framing effect, then they would be changing without changes to any relevant moral facts, and so they would not be reliable.

A defender of the use of intuitions may object that these are tests performed on laypeople who are not experts in moral philosophy. As experts in the use of intuitions and moral philosophy, one might argue, the intuitions of the moral philosopher are more likely to be reliable and less vulnerable to the framing effect than those of the laypeople tested by psychologists. Although I defend a limited account of moral expertise in chapter 2, the notion of expertise in the formation of intuitions is much harder to defend. Most problematically, the claim that the intuitions of moral philosophers are less vulnerable to the framing effect is an empirical question, and cannot be established without appealing to evidence showing it to be the case. Unfortunately for the intuitionist, the evidence that has been collected does not seem to support this claim. The research that has been conducted thus far seems to indicate that philosophers are not significantly less likely to be influenced by framing or ordering effects than non-philosophers.<sup>24</sup> So this avenue of argument is unlikely to be successful.

The question that remains, then, is whether or not our moral intuitions are vulnerable to the framing effect. As we have already seen, the first example of the framing effect is one in which it is affecting moral intuitions, but this alone is not necessarily enough to cause us to assent to the claim that our moral intuitions are subject to the framing effect in many circumstances. It may be that there is something unique or

---

<sup>23</sup> i.e., they should not change unless some morally relevant factor changes.

<sup>24</sup> Cf. Schwitzgebel and Cushman 2012; Tobia, Buckwalter, and Stich 2013; Schwitzgebel and Cushman 2015

odd about these probability-based cases,<sup>25</sup> or something else unusual about this that will not occur for our other moral intuitions. Unfortunately, however, this does not appear to be the case, as other studies have found a similar vulnerability to framing effects among broader sets of moral intuitions.

Petrinovich and O'Neill (1996) tested for framing effects using the trolley problem. They presented test subjects with variants of the trolley problem. In half of the cases, the problem was framed as a question of how many people would be saved,<sup>26</sup> while the other half are presented the problem in terms of the number of people killed<sup>27</sup> (ibid. 149). Participants were then asked to rank how strongly they agreed or disagreed with the statements that one should or should not flip the switch. They found that for both flipping the switch and refusing to flip the switch, "the Save wording resulted in a greater likelihood that people would absolutely agree, and the level of agreement was stronger" (ibid. 152). In other words, the group who were presented with the problem in terms of the number of people to be saved were more likely to agree – and agree more strongly – with both flipping and not flipping the switch, while those who considered the question in terms of how many people would be killed were more likely to disagree with either option. As with Tversky and Kahneman's experiment, nothing of any plausible moral relevance was changed between the two test groups, and yet they had differing intuitions. So, in this case as well, our intuitions are subject to the framing effect. On top of this, when Petrinovich and O'Neill presented subjects with three different versions of the trolley problem,<sup>28</sup> their intuitions about the cases varied depending on the order in which

---

<sup>25</sup> Psychologists have found that people are surprisingly bad at estimating probability, or making any intuitive judgments about it. A prime example of this is the base rate fallacy, which sees people consistently ignore information about low base rates in favour of information about a particular case (cf. Bar-Hillel 1980, Tversky & Kahneman 1974).

<sup>26</sup> i.e., flipping the switch saves five, not flipping the switch saves one

<sup>27</sup> i.e., flipping the switch will cause one person to die, not doing so will cause five people to die

<sup>28</sup> The standard switch-flipping version, pushing the fat man, and pushing a button causing the trolley to run over a bystander on a footbridge



they were presented (ibid. 156-8).<sup>29</sup> Again, as moral philosophers we should want to say that the order in which we consider possible moral dilemmas makes no difference to their moral quality. If it is wrong to push the fat man off the footbridge, it is just as wrong whether we consider it before or after considering a “standard” trolley case. Once again, then, we see moral intuitions being vulnerable to framing effects.

Although the evidence in Petrinovich and O’Neill’s experiments is not significantly different from Tversky and Kahneman’s – intuitions of the kind commonly employed by moral philosophers being vulnerable to the framing effect – seeing the ways in which intuitions around trolley cases can be altered by the framing effect should cause us significant concern. I stated earlier that we may not necessarily be too concerned with the result of Tversky and Kahneman’s experiment because there may be something unusual about that particular intuition that causes it to be vulnerable to framing effects in ways other moral intuitions are not. It is not so easy to claim this about trolley cases, however. The trolley problem is a fairly typical example in moral philosophy, and one which is heavily employed. It is also used to illustrate/justify the distinction between killing and letting die, and our differing intuitions about some variants of the trolley case are used to support the moral importance of double-effect, or to criticize some consequentialist positions by showing that we think it is acceptable to sacrifice one life to save five in some trolley cases but not others. If the reliability of this intuition is called into question, then we must also question how much we can trust the principles derived from it, some of which are central moral principles.

The only conclusion we can draw, then, is that our moral intuitions are subject to the framing effect in some circumstances, and thus unreliable. Since moral philosophy as traditionally approached cannot proceed without making use of intuitions, this presents a challenge. To develop and advance our theories, we need a reliable base on intuition and first principles to build from. But if we have reason to doubt the reliability of our

---

<sup>29</sup> This is sometimes referred to as the order effect, or order effect bias, and is frequently taken to be a form of the framing effect.

intuitions, then it is difficult to see how we can proceed except on a highly provisional basis.<sup>30</sup>

Although problematic, this is not necessarily a devastating blow to non-empirical approaches to moral philosophy. As Shafer-Landau (2008) points out in his criticism of *the master argument*, there is a big difference between the claim that some (or even many) of our intuitions are vulnerable to framing effects and claiming that all of them are vulnerable to framing effects. Similarly, he points out that even if we grant that our intuitions are vulnerable to framing effects in many circumstances, it is not clear which of two propositions we may be assenting to. We may mean that a small number of our intuitions are vulnerable to the effect in a large number of circumstances, or that many or most of our intuitions are vulnerable to it (ibid. 86). All that *the master argument* – combined with the empirical data we have so far gathered – shows is the former possibilities of each of those two. And so, one might object, it may well be that many of our most important or foundational intuitions are not vulnerable to framing effect. Indeed, it is difficult to imagine that, simply through changing how the question is posed, I could come to reject my intuition that torturing innocent babies merely for the fun of doing it is wrong.

Although the vulnerability of intuitions to framing effects is not a devastating problem for traditional approaches to ethics, it is nevertheless a significant enough issue to warrant concern. A point Shafer-Landau makes in defense of intuitionism actually serves as a perfect statement of exactly why we should be concerned, even if we only have evidence that some of our intuitions are vulnerable to framing effects. He claims that the empirical studies only reveal “that some such processes are unreliable” (ibid., 88), and that, “I don’t know whose processes, or which processes, are likelier than not to lead to error” (89). And while that is true – we cannot point to any given intuition and say definitively that it is or is not vulnerable to the framing effect, and thus formed unreliably – that is also exactly the problem.

---

<sup>30</sup> e.g., “If this intuition is correct, then this broader set of principles follows.”

To see exactly why this uncertainty is a concern, let us consider an analogy to perception. I know that, under ordinary circumstances, my senses are more or less reliable, even if they are subject to error under some circumstances of which I am aware. And so, when I have no reason to think any of those circumstances apply, I can be confident that I am justified in judgments I make based on data from my senses, even though there are potential undermining factors or defeaters for those judgments.<sup>31</sup> Now imagine I am in the grip of a severe fever that is causing me to hallucinate vividly. I am not always hallucinating, and indeed most of the time, I am perceiving my surroundings exactly as they are.<sup>32</sup> To further complicate matters, as a result of both the fever and the vividness of the hallucinations, I cannot reliably tell whether I am hallucinating or lucid at any particular moment. In this case, even though I may form a number of beliefs about my surroundings that are entirely correct, and formed through an ordinary, reliable process, I should nevertheless be concerned that any of the judgments I make are not in fact justified, since I cannot tell the difference between the impressions that were formed by a reliable process and those that were not. In this case, I would need some other way to test my judgments to know which of my perceptions are accurate, and which are not.

The problem with our current epistemic state regarding our moral intuitions is that it is closer to the fever case than the ordinary case. There are clear cases in which our intuitions are vulnerable to framing effects, and although in many – or even most – other cases, it may turn out that our intuitions are perfectly reliable, we are not able to tell which intuitions fall into which category. As a result of this, even though many of our intuitions may be reliable and justified, we are not in a position to know which ones are, and thus can offer no evidence that we actually have the reliability and justification that would be necessary to be able to rely on our intuitions. In short, the framing effect, like a high fever, gives us a reason to be skeptical of our intuitions.

---

<sup>31</sup> Absent certain forms of skepticism.

<sup>32</sup> Or, at least as exact as I am capable of being under ordinary circumstances.

There is, of course, a potential way to resolve this issue. The philosopher who wishes to rely on intuitions may point out that all she needs to do to rescue intuitions is determine which intuitions are and aren't vulnerable to framing effects, and the circumstances and ways in which framing effects can manipulate our intuitions. Then we will know which intuitions are reliable, and which cases we need to be worried about. With this information, we could presumably consider any particular intuition we wish to use as evidence and know with some level of certainty whether it is a reliable intuition, or one which is likely to have been formed as the result of some sort of framing effect or other cognitive bias. If the intuition is reliable, then it is fine to use as evidence. If, however, the intuition is such that it is likely to have been influenced by some cognitive bias or other, then that is a sign that we should look for further evidence to support that intuition before drawing any significant conclusions from it.<sup>33</sup>

Having such complete information would indeed rescue the use of intuitions in moral philosophy. The main reason we don't know exactly which intuitions are vulnerable to framing effects or other cognitive biases, or the exact circumstances under which our intuitions are able to be manipulated is that there is still much more research to be done. It is also the case, however, that this is exactly the position laid out by Knobe and Nichols that I referenced at the outset of this chapter: to be able to effectively do moral (or any) philosophy, we must be aware of how facts about our psychological makeup shape and inform our intuitions and judgments. Our concern here is not with whether or not the empirical approach to ethics is able to deal with these challenges. What is at issue for this discussion is whether or not the reliance on intuitions which is at the methodological core of the traditional, non-empirical approach to ethics can be maintained in the face of challenges from cognitive science and sociobiology. To answer

---

<sup>33</sup> It should be pointed out that even this is a highly idealized scenario in which a person is able to have fairly thorough knowledge about the sources and causes of her individual beliefs, and hold onto or discard those beliefs in a way that anyone who isn't perfectly rational is ever likely to be able to do. In an actual situation, we might expect any number of possible reactions from someone who has learned that her intuitions are likely to be the result of framing effects. These reactions could include revising her beliefs, stubbornly digging in and insisting that the intuition is correct despite possible problematic influences, or attempting to rationalize or explain away the apparent problem in the way Jonathan Haidt has explored (Haidt 2001). For our purposes right now, however, this idealized version will suffice.

this question by appealing to the ability of psychologists or experimental philosophers to clarify our concepts and use of intuitions is to admit that the traditional methods of ethics are inadequate to establish the necessary robustness of intuitions.

The traditional approach to moral philosophy cannot do without the use of intuitions on at least some level, but this use of intuitions is only methodologically sound if we can know that our intuitions are reliable. As the framing effect shows, however, at least some of our intuitions are subject to manipulation by seemingly irrelevant factors. Since the reliability of moral intuitions depends in part on those intuitions only changing for morally relevant reasons, we cannot treat intuitions vulnerable to the framing effect as reliable. On top of this are two further complications: First, the framing effect is just one of a myriad of cognitive biases that could potentially interfere with the reliability of our intuitions. Second, there is no a priori way to determine which of our moral intuitions are vulnerable to these effects. Considering all of this together, we must come to the conclusion that traditional approaches to ethics face defeaters which ought to undermine our confidence in the reliability of our moral intuitions.

### 1.3 The Darwinian Dilemma

A different challenge to the reliability of intuitions in moral philosophy comes from a dilemma raised by Street (2006). She points to a tension between our knowledge that we are evolved creatures and realist accounts of value. She argues that many of our evaluative judgments have been strongly influenced by Darwinian selection pressures (113-21). It is important to note that she is not arguing that the sorts of explicit, considered beliefs like, “We ought to avoid doing needless harm to others” are direct results of natural selection. Rather, she argues that our more basic evaluative tendencies, “an unreflective, non-linguistic, motivational tendency to experience something as ‘called for’ or ‘demanded’ in itself, or to experience one thing as ‘calling for’ or ‘counting in favor of’ something else” (119) are directly influenced by selection pressures, and these basic tendencies have strong influences on our fully-formed judgments (119-20). She then presents two options for a realist to explain the relationship between the evolutionary influences on our moral judgment and the independent moral truth of realism. The realist can claim that there is no relation between the two and that the

influence of natural selection on our judgments is “purely distorting” (121), or they can say there is a relation in the form of those evolutionary factors “as having *tracked* the truth” (125, original emphasis). As the details of Street’s response to the latter position are outside the scope of the current discussion, I will focus on the former response.

Street’s main argument against the response that there is no connection between the Darwinian forces shaping moral judgments is that it leads to the skeptical conclusion that most of our moral judgments are probably mistaken. If no connection exists, then the influence of selection pressures on our judgments distorts them (rendering them unreliable) and at best only moves us towards true judgments by chance (122). This is a similar argument to the one we just discussed concerning the ways in which the framing effect makes our intuitions unreliable. If these judgments are formed by a process which is disconnected from their truth conditions, they only manage to arrive at the right answer by chance, not by being a reliable guide to the truth we seek. And so, Street concludes, this response leads us to an implausible moral skepticism (*ibid.*).

Street also considers the possible rejoinder that, although the content of our judgments is likely to be distorted, that is not a particular problem, because our ability to rationally reflect on and revise our judgments can correct for the distorting influences of selection pressures (123). This point is also relevant to the earlier discussion on the framing effect, as it is also a possible response to the issues I raised. We could accept that the framing effect can manipulate our intuitions, but argue that our ability to reflect on these intuitions gives us the ability to find and correct distortions from framing effects or other cognitive biases, thus smoothing out any potential wrinkles in our intuitions. Unfortunately, this argument will not work in either case. As Street points out, it is a mistake to think – especially in the context of moral judgment – that rational reflection is an “uncontaminated tool” (124). Rather, we can evaluate our moral judgments only by comparing them against our other moral judgments, seeking justification only from other moral claims (*ibid.*). In this case, though, our attempts to smooth out the distortions from selection pressures or framing effects in our moral judgments and intuitions through rational reflection are exactly as problematic as those judgments and intuitions were in

the first place. Once again, we see evidence about the sources of our intuitions undermining our confidence in their reliability.

There is another potential response to this dilemma that Street does not consider: rather than attempting to reconcile evolutionary pressures on moral judgments with moral realism, or rejecting moral realism – as is Street’s preferred option – one could instead reject the first premise and claim that our moral judgments are not the result of Darwinian selection pressures. Considering this option will move us into a minor digression, as it does not strictly concern the reliability of our moral intuitions. However, the problems with such a response are illustrative of more general problems for traditional approaches, so it is worth pausing to examine now.

The most extreme version of this position is put forth by Thomas Nagel in *Mind and Cosmos* (2012). He agrees with Street’s argument that moral realism is incompatible with our understanding of evolution by natural selection, but argues that, “since moral realism is true, a Darwinian account of the motives underlying moral judgment must be false, in spite of the scientific consensus in its favor” (105).<sup>34</sup> He is committed to this idea to the point of claiming that “the scientific credentials of Darwinism [. . .] are not enough to dislodge the immediate conviction that objectivity is not an illusion with respect to basic judgments of value” (110). Nagel’s argument for this is practically non-existent, consisting mainly in an assertion that he cannot help but think of pain as being truly bad and pleasure as truly good, no matter how hard he tries to imagine that this is not the case (ibid.). The details of his argument are not particularly relevant, however. What is of interest is where this notion goes wrong.

It should seem immediately implausible that, upon acknowledging a tension between our moral intuitions and a Darwinian understanding of life, the correct response could possibly be to decide to hold on to our intuitions and jettison natural selection. Indeed, in this case, I would agree that the (I hope) common intuition that Nagel’s

---

<sup>34</sup> This position is mistaken almost to the point of absurdity, and does a disservice to other non-empirical accounts of morality by being so extreme as to make other more reasonable positions look untenable. Despite this, it is informative in the ways it goes wrong, and so is worth looking at more closely.

judgment has gone seriously awry is the correct one. Nevertheless, it is worth asking exactly why this is mistaken.

The first thing to notice is that it is not an incoherent position to take. If beliefs *x* and *y* are in conflict with each other, absent further information, we can resolve that conflict by rejecting either one of them, and thus have a consistent belief set. As a very basic epistemic principle, however, when faced with the choice of which belief to reject, and without conclusive evidence about which is false, we should reject the one that is less likely to be true.<sup>35</sup> In this case, the two beliefs we must choose between are (a) my basic moral intuitions accurately reflect reality, and (b) evolutionary forces have shaped our intuitions. But in order to reject (b), we must reject a more general Darwinian claim, (b\*) the theory of evolution by natural selection accurately describes how living things, including humans, have come to be. The question becomes, which of (a) and (b\*) is more likely to be true? If it is not clear which is the case, it can be helpful to think about what would need to be the case for either claim to be false. For (a) to be false, my moral intuitions would need to have been led astray by any number of possible factors, including the sorts of cognitive biases discussed above, but also the various moral skepticism arguments that have been advanced throughout the history of philosophy. For (b\*) to be false, the entire modern study of biology, and everything built on that understanding must be mistaken. In other words, an entire field of science, and everyone working in that field must be colossally off-base. So, while there are many ways in which (a) can be false, quite a lot would have to be wrong for (b\*) – and thus (b) – to be false, and so (a) seems like the less likely belief. Ultimately, Nagel's mistake is in not fully appreciating just what it would mean for the Darwinian worldview to be deeply mistaken in the way he would need it to be in order to simply reject it out of hand.

This may have seemed like a lengthy digression, but lessons from it can be applied more broadly to attempts to salvage moral intuitions from apparent empirical defeaters. A large part of the reason (b\*) is much more likely to be true than (a) is that

---

<sup>35</sup> I take this to be an uncontroversial principle



scientific claims are typically supported by a variety of data, while moral intuitions are notoriously difficult to justify. This has broader implications for any response to a science-based critique of moral philosophy. In particular, it shows the need for caution in dismissing science-based challenges. For instance, imagine that one were to respond by stating that, although they see the evidence that the framing effect exists, we should nevertheless continue to treat our moral intuitions as unproblematic. Or, similarly, a philosopher who simply ignores any challenges from empirical sources. In both these cases, what is being tacitly argued is that our confidence in our moral intuitions is (or should be) sufficiently high that we can be more confident in them than a well-supported scientific theory. This is not as dramatic a form of the mistake as Nagel makes, but it is nevertheless highly implausible to think a priori that it is more likely that the research on the framing effect has all been somehow mistaken than it is that our moral intuitions are potentially biased in ways that should concern us.

This does not mean that there can be no reasonable response to these sorts of studies from traditional moral philosophers. As I will argue in the next chapter, there are legitimate grounds on which to criticize the assumptions of the empirical approach, including casting doubt on some of the studies used to support that position. It does, however, restrict how proponents of the traditional approach can respond. Continuing to use the reliability of intuitions as our example, there are two main ways to recover the reliability of intuitions. One is to gather further data that show that the distorting effects of things like cognitive biases or selection pressures are not as strong as we had suspected or feared, and that most of our most important intuitions are generally reliable. As stated above, however, this is an empirical task, and not one suited to non-empirical approaches to ethics. The alternative is to argue not that the results are incorrect, or to treat our intuitions as being more justified or reliable, but rather to argue that the data do not apply to the sorts of situations we care about, or do not apply broadly enough to be a concern. A prime way to do this is by arguing that the conclusions drawn by the scientist or philosopher presenting the study has reached beyond what is warranted by the data, or that alternate explanations – ones not as problematic for the view being defended – that

have not been adequately considered or dealt with. Traditional philosophers must, then, engage directly with empirical challenges to moral philosophy.<sup>36</sup> Unfortunately, such responses will not help to explain away the distorting effects of cognitive biases or selection pressures. Potential sources of influence on our moral intuitions are undeniably relevant to their reliability, and as argued above, while we do not yet know the full scope of the framing effect, that is exactly the problem, as we cannot know if any given intuition of ours has been distorted by it. Although it will not be successful for saving the reliability of intuitions, this is a useful method for traditional moral philosophers to push back against empirical criticisms.

## 1.4 Thought Experiments and Moral Counterexamples

The other central investigative tool of philosophers – and especially moral philosophers – with which I am interested for this chapter is the thought experiment. Contemporary moral philosophy is replete with examples of famous and influential thought experiments like the trolley problem or Thomson’s violinist, which most philosophers can at least give an outline of without needing to think about it. This section is concerned with how thought experiments are used in moral philosophy, and especially how they make use of intuitions. I argue that some common uses of thought experiments reveal a confidence in the reliability of intuitions which cannot be adequately supported even from within the traditional moral philosophical framework, and certainly cannot handle the sorts of skeptical worries discussed above. In light of this, I conclude that the reliance on certain forms of thought experiments poses a methodological problem for traditional moral philosophers.

Although common, thought experiments are not treated uncritically by philosophers. Common objections to the use of thought experiments exist, some focusing only on particular sorts of thought experiments, and others arguing against the usefulness of thought experiments in general – and especially in ethics. To the former case, some argue that particularly outlandish thought experiments – that is to say, thought

---

<sup>36</sup> I discuss exactly how this can be done in chapter 4.

experiments that involve situations that are far removed from any sort of ordinary experience we could possibly have – are suspect because we cannot form reliable intuitions about them.<sup>37</sup> Daniel Dennett’s complaint about “intuition pumps” in thought experiments which serve not as “an engine of discovery, but a persuader or pedagogical tool” (Dennett 1980, 429) is also a complaint along these lines: it is not an attack against thought experiments in general, but particular uses of them. Attacking the use of thought experiments more generally, moral particularists such as Jonathan Dancy (1985) have argued that thought experiments are necessarily underdescribed and leave out vital details that are necessary to any proper moral judgment. Despite these objections, however, thought experiments continue to be widely used and are a generally accepted tool in moral philosophy. Given this, it is worth considering their use as a methodological tool in moral philosophy.

Thought experiments can be broken into two general types: constructive thought experiments, which seek to support a theory or prove that something is true, and destructive thought experiments, which serve to undermine or provide a counterexample to an existing theory.<sup>38</sup> Within the set of constructive moral thought experiments, we can divide things into two further classes that I will call the clarifying, and the supporting. Clarifying thought experiments do not so much seek to argue for a position as to help clarify what the position entails, or how it applies. Kant’s use of the murderer at the door example (Kant 427) would fall into this category. He is not trying to convince his reader that they cannot be morally justified in lying to the murderer. Instead, Kant is attempting to show how his system of ethics understands this situation. The supporting thought experiment seeks to establish some moral fact or principle. The various uses of the trolley problem that have served to draw out the distinction between killing and letting die and establish that this distinction is meaningful would be examples of this.

---

<sup>37</sup> Kasper Lippert-Rasmussen (2008), for instance, presents such an argument.

<sup>38</sup> My use of these terms is taken from Brown 2004. Brown also recognizes a third category, the Platonic thought experiment, but these are beyond the scope of this investigation. I should also note that I am not necessarily taking any of these distinctions to represent distinct natural kinds. Instead, these classifications should be thought of as useful ways that we can talk about the different uses of thought experiments

Similarly, destructive thought experiments can also be broken into two further subtypes. Let us call these the conflicting thought experiments and countering. The former point out a conflict between our beliefs, often between two intuitions, or a philosophical position and pre-philosophical belief. Thomson's Violinist (1971) is an instance of this type of thought experiment. Showing that there is a conflict between the intuition that abortion is always morally impermissible and the intuition that unplugging oneself from the violinist is permissible. Since these two are in conflict, one must be rejected, and Thomson then provides an argument to support rejection of the belief that abortion is impermissible. The final class seeks to prove that a moral theory is wrong by presenting a case which has all the qualities of a morally permissible action according to that theory, but which is nevertheless morally impermissible (or vice versa). One of the more famous moral counterexamples is Nozick's experience machine (Nozick 1974, 42-3) meant to undermine the utilitarian's claim that experience is all that matters by presenting a situation in which we care about something beyond just experience.

The difference between these two may not be immediately clear, so it will be worth going over. Initially, they both seem to work the same way: prove that something we think is true conflicts with our intuitions, and thus should be discarded. The main difference will generally be in what they take as their target. Conflicting thought experiments point to a contradiction between two simultaneously-held intuitions, and are generally going to concern themselves with pre-philosophical notions. Moral counterexamples, on the other hand, take established philosophical theories as their target. So, after Philosopher X has drawn from intuitions, rational reflection, reflective equilibrium and so forth to build up some moral theory, Philosopher Y responds by pointing out that some action or other has all the characteristics that matter to X's theory, but would be counterintuitive to claim it is in fact morally permissible, in the non-moral realm, Gettier cases are the most popular examples of this, taking the established philosophical account of knowledge as justified true belief, then pointing to a situation in which something that counts as a justified true belief but is intuitively not knowledge. In short, conflicting thought experiments focus on pre-philosophical views while moral counterexamples take on argued-for theories.

Regardless of our views about intuitions and whether or not they are reliable, clarifying and conflicting thought experiments would seem to be acceptable and useful investigative tools. A clarifying thought experiment is a pedagogical tool more than an argumentative one, and if questionable intuitions are involved, this would be as part of the construction of the theory itself, not the example that seeks to spell out what it says. Similarly, although intuitions are necessarily involved in pointing out that two intuitions are conflicting, it is hard to see where we might object on the grounds that the intuitions involved are unreliable. All that is really being argued in such a thought experiment is that at least one of the intuitions needs to be changed or discarded, given the epistemic requirement that we attempt to maintain a consistent set of beliefs. Even if both of the intuitions are formed under epistemically suspect circumstances, we must nevertheless think that rationality compels us to resolve the inconsistency between them even as we remain skeptical regarding the truth of either.

This leaves us with supporting thought experiments and moral counterexamples. Both of these will prove problematic. Supporting thought experiments run into exactly the problems discussed above: they elicit intuitions about what is morally permissible or impermissible in particular cases, then argue for general principles derived from those intuitions. But as I have already argued, these intuitions are unreliable, and so we cannot know if we are justified in drawing inferences from them, and so this use of thought experiments is epistemically dubious. If we cannot rely on the intuition, then the evidential power of a thought experiment employing the intuition must be diminished. Finally, then, we come to moral counterexamples. We could criticize them for their own use of intuitions, given the problems already discussed, but I want to set those concerns aside for now. Instead, I want to argue that, even from within the point of view of the traditionalist, these sorts of thought experiments ought to be treated with suspicion, as they overstate the usefulness of intuition.

To this end, let us consider Nozick's experience machine as a case study.<sup>39</sup> In his argument against hedonic theories of value (and thus against certain forms of utilitarianism), the reader is asked to imagine a machine. This machine can be programmed to stimulate your brain in order to give you any experience you may want to have. The experiences given by the machine will be perfectly indistinguishable from experiences outside of the machine, and once you have plugged in, you will not remember doing so. Researchers have thoroughly and broadly investigated the experiences of other people, so there is no chance of you missing out on any desirable experiences, as they are all on the list of experiences you can choose from the machine, and every two years you can leave the machine for ten minutes to program the next two years of your experiences (and again, once plugged back in you forget that you were ever out), so your experiences in the machine will be exactly what you would want your life to be like, were you given a choice. While you are having all of these experiences, however, you are actually just floating in a large tank with electrodes strapped to you to stimulate your brain in the correct way. The result of the thought experiment, according to Nozick, is that we would choose not to plug into the machine, which leads him to conclude that experience is not all that matters (Nozick 1974, 42-3). Nozick also adds additional considerations and abilities to the machine, but for the purposes of this discussion, these are the only aspects of the thought experiment with which we need to be concerned.

The particular target of this thought experiment is the notion that ultimately our experiences are all that matter. For the sake of argument, let us grant that Nozick's central claim is correct, and most – if not all – of us would choose not to plug into the machine, and further that the reason we would choose this is that we think experience is not all that matters. That is to say, we agree with Nozick's claim that, “we want to *do* certain things, and not just have the experience of having done them” (ibid., 43). Let us also suppose that, “our experiences are all that matter” is a claim that his opponents are

---

<sup>39</sup> It will be important to remember throughout this discussion that I am considering the thought experiment in isolation, not the thought experiment in conjunction with the rest of Nozick's argument, let alone the rest of *Anarchy, State, and Utopia*.

genuinely committed to.<sup>40</sup> So, his belief about our intuitions is correct, as is his understanding of his philosophical target. The question I want to address now is, what are we actually justified in concluding from this thought experiment, even if we assume that Nozick is right about these two central ideas?

There are two important things to bear in mind when determining what we should conclude from the experience machine. First, the notion that experience is all that matters is not an intuition or premise of certain hedonic forms of utilitarianism. What it is instead is the conclusion of a series of arguments all intended to establish the point that experience is all that matters. That is to say, it is a point that has been proven by exploring and probing intuitions, analyzing the consequences of those intuitions, and so on. The second thing to bear in mind is that our response that we would not plug into the experience machine is not itself an argument that it is wrong to do so. By itself, and absent any argument for why this reaction is correct, it is just an intuition. Bearing both of these in mind, it becomes easy to see where this thought experiment goes wrong.

Although moral philosophy cannot proceed without the use of intuitions, the task of moral philosophy is not simply to list or conform to all of our moral intuitions. Moral philosophy is a normative enterprise, concerned with what we *should* do, not with what our intuitions tell us we should do.<sup>41</sup> If this is the case, then it must also be true that, in at least some cases, what is morally right will diverge from what our intuitions tell us is right (absent an extreme amount of luck on the part of our intuitions). Exactly when and how we should jettison an intuition – even one which seems clear or obvious – is a central question to moral philosophy, and not one I will be able to answer here, but presumably one of the reasons to reject an intuition is that it is incompatible with principles that have been established through argumentation. Simply put, if an argument is sound, then pointing out that its consequence is counterintuitive is irrelevant.

---

<sup>40</sup> It is not entirely clear to me that even the classical utilitarians would completely agree with this premise (at least not without qualifications), but for the purposes of this discussion, we can set this concern aside and grant that this is a consequence of hedonic theories of value.

<sup>41</sup> This point is defended further in chapter 2.

This is perhaps easier to see if we consider a non-moral example. Suppose that everyone found the idea that there are some mathematical statements which are undecidable, or that it is impossible for a formal system to prove itself consistent to be wildly counterintuitive. If the conclusion we came to from this was that Gödel's incompleteness theorems should be rejected, our reasoning would have gone horribly wrong. If we want to reject the theorems, we would have to do it by showing that something has gone wrong with the proofs. Similarly, if the arguments of the utilitarian lead to the conclusion that experiences are all that matter, then an intuition that, in one case, there appears to be something other than experience that is relevant should not be enough to undermine the conclusion. To reject the conclusion, we would have to show what has gone wrong with the actual argument; either a false premise, or an unjustified inference at some point. The experience machine thought experiment does not offer this. Instead, it offers the intuition that something other than experience matters, and holds that up against what amounts to a series of arguments that only experience matters. If this problem sounds familiar, it should, as it is exactly the issue raised by Nagel's response to Street. The experience machine presents us with two conflicting beliefs, one of which is grounded only by intuition, the second of which is established by evidence and argument. As before, we must reject one, but also as before, we are asked to reject the one which is less supported, and thus less likely to be true. Thus, the experience machine thought experiment fails to adequately weigh the evidence on both sides of the debate, and asks us to make the dubious decision to reject the conclusion of a series of philosophical arguments, rather than rejecting the intuition.

I have been focusing on the experience machine up to now in this section, but this challenge will apply to any attempted moral counterexample which seeks to use a moral intuition to undermine a position reached by philosophical reflection and argument. Such thought experiments can only show that the theories in question are counterintuitive, but unless we hold the implausible idea that moral theories must adhere perfectly to our intuitions, this is a very different thing than showing that the theory is false. Absent some independent reason to think the argument supporting theory is flawed, it will always be more likely that the intuition is mistaken, and thus that the intuition is the belief that should be discarded or revised.



So, even without bringing in the concerns that empirical data give us about the reliability of our conclusions, we see significant challenges posed to one of the major tools of moral argumentation. That is to say, even from within the traditional approach to moral philosophy, moral counterexamples are a problematic methodological tool. If we also add in the concerns discussed earlier about the reliability of our intuitions, and various distorting factors and cognitive biases that interfere with them, the use of a moral intuition as an undermining factor for an established theory becomes even more epistemically problematic. On top of this, our inability to rely on our intuitions also makes thought experiments less useful as a tool to support our theories, as we must see the evidence provided by them as weaker and more prone to error. All of this means that two of the major uses of thought experiments employed by moral philosophers are not able to provide as strong evidence as we would like.

This is not necessarily a decisive argument against the use of thought experiments of thought experiments in moral philosophy. As I said at the outset of this section, two of the major uses of thought experiments are almost entirely unproblematic, and thus do not need to be reconsidered. Supporting and countering thought experiments face problems, but this does not necessarily mean they cannot or should not be used. All it really means is that we need to be aware of what kind of evidence they provide, and how strong that evidence is. The problem arises when these thought experiments – and especially moral counterexamples – are treated as decisive evidence, rather than as some (possibly minimal) reason to re-examine our moral ideas, or as a beginning point for an argument. They can supplement further arguments, but cannot serve as arguments by themselves. The use of thought experiments in moral philosophy is not unproblematic, and they cannot be used without careful attention to their evidential value, but they need not be discarded outright.

## 1.5 Conclusion

Bringing all of this together, we must come to the conclusion that moral intuitions are not sufficiently reliable to be able to serve as a basis for moral reasoning. The non-empirical approach to ethics must rely on intuitions to avoid an infinite regress of inferential justification, but this can only proceed on the assumption that these intuitions are at least

somewhat robust and reliable. To be reliable, our intuitions about moral cases should remain fixed, so long as nothing of any moral relevance has changed. Similarly, distorting influences that cannot be explained away also affect the reliability of our intuitions. Empirical research from psychology, cognitive science, and sociobiology shows us that there are a number of apparently morally-irrelevant factors that can influence our intuitions. The prevalence of cognitive biases such as the framing effect shows us that our intuitions can be manipulated by factors as seemingly irrelevant as the order in which the options are presented. At the same time, even were our intuitions free of these effects, the sorts of selection pressures that have influenced us throughout our evolutionary history are highly unlikely to have geared our most basic moral intuitions toward truth tracking. More likely, our moral intuitions are influenced by evolutionary fitness, acting as a distorting effect on our intuitions and making them less likely to be true. This also causes issues for philosophical tools that rely on the use of intuitions. At the very least, we must be cautious of two of the four major ways though experiments can be used in moral philosophy, since they rely heavily on moral intuitions, and thus need further support or confirmation. Taken together, all of this means that the traditional approach to moral philosophy needs to be regarded with at least some degree of skepticism owing to the issues raised by proponents of empirical and experimental philosophy.

## Chapter 2

### 2 Methodological Assumptions of an Empirical Approach to Ethics

If – as was argued in the previous chapter – the empirical challenge poses serious problems for traditional approaches to moral philosophy, this would seem like good news for the empirical approach. If traditional approaches face problems arising from empirical data, an approach that incorporates empirical data would seem to be able to solve this problem, or at least presents a means of approaching the problems. This makes the empirical approach a much more attractive possibility, since it is necessarily responsive to empirical data. In order for this approach to be viable, however, it must itself be able to address any significant methodological challenges that may be raised against it. This chapter will consider the approaches empirically-minded philosophers take to moral philosophy, and the methodological assumptions about ethics which underlie these approaches. There are, therefore, two central questions for this chapter. First, what must be true about moral philosophy or moral claims for an empirical approach to ethics to be plausible? Second, following from this first question, should we think that this is the correct way to understand moral philosophy?

This chapter has four sections. In the first, I identify three major assumptions made by empirically minded ethicists in order to motivate their projects. The following three sections will examine each of these assumptions in turn, arguing that we have reason to be skeptical about them. In the second section, I challenge the assumption that the work of moral philosophers examining moral questions and dilemmas does not provide them any expertise or special insight into moral problems. The third section examines the assumption that ethics is a descriptive project, rather than a normative one, and argues that this does not accurately capture the practice of moral philosophers. In the final section, I argue that moral theories do not have the sort of empirical consequences which makes them testable or falsifiable through scientific inquiry. I conclude that with the assumptions underlying the empirical project lacking adequate justification, we ought to be highly skeptical of empirical approaches to moral philosophy.

## 2.1 The Empirical Assumptions

There are several different ways of gathering data for an empirical inquiry into ethics, and several different types of data that could be used. This can be as small a difference as the choice by a philosopher to conduct experiments or surveys herself – as many who work in the tradition of experimental philosophy do, often designing and conducting their own experiments, as will be discussed below – rather than gather data from studies conducted by others, and analyze and interpret the data to see what sorts of conclusions about morality may be drawn. Jesse Prinz looks to cross-cultural studies to justify his account of moral judgment, and particularly his agent relativist view that moral judgments are based on emotion (Prinz 2007a, 2007b, 2008, 2009). Patricia Churchland prefers to ground her moral inquiry in neuroscience and evolution, discussing the aspects of the brain which serve as the platform for morality (Churchland 2011). Shaun Nichols has pointed to various studies involving psychopaths to defend his motivational externalism and attempt to undermine internalism (Nichols 2002, 2008). These are varied approaches, but they share a common theme: the central use of data from psychology, neuroscience, and evolutionary biology to defend and justify normative moral claims.

There are three major background methodological assumptions made by philosophers attempting to derive rich normative or metaethical results from empirical data, some subset of which must be held for the project to be plausible. First, the moral intuitions and judgments of philosophers are not to be privileged over those of non-philosophers. Philosophers do not possess any expertise which would make their intuitions about moral matters more robust, reliable, or free from bias than those of anyone else. Put simply, there is no expertise in moral matters. The second is a belief about what the project of moral philosophy is at a fundamental level. Specifically, it concerns the belief that the project of ethics is a descriptive one. What we are aiming to do as moral philosophers is accurately describe the norms people hold, or find a unifying essence that is the cause of our moral judgments.<sup>42</sup> The third concerns the consequences

---

<sup>42</sup> That is, some single thing that our moral judgments have in common, which can be understood to be the cause of those judgments

of moral theories. It holds that at least some of the consequences of moral theories will hold that people will feel, act, or reason in certain ways, and we can judge the truth of the theory by seeing whether or not these consequences actually occur. In other words, moral theories have empirical “results” and can be falsified by showing that those results do not occur in the world.

There are three things to note about this set of assumptions. First, they are compatible with each other, but independent. Any subset of the three beliefs may coherently be held. Second, any one of these assumptions is sufficient for justifying empirical inquiry into moral questions. In the first case, if philosophers have no special insight into morality, we can gain greater insight by expanding the number and kinds of people whose intuitions we are sampling. In the second, if the goal of moral philosophy is to accurately describe people’s actual norms – what people actually think is right or wrong – then we can do this by testing people to see what norms they actually hold. In the third, if a theory has as a consequence that people will, for example, prefer option a to option b, and we can empirically determine that people in fact prefer option b, then we have shown that the theory gets something wrong, and we can reject it. Finally, while no particular one of these beliefs is necessary to ground an empirical project, it is necessary for empirical ethicists to accept at least one of them. If we believe that the project of ethics is a normative project concerned with what we ought to value rather than what we do, that philosophers have unique and privileged insight into morality, and that there aren’t testable results to moral claims, then it is not clear what – if anything – could plausibly be illuminated by empirical inquiry, as we now seem to have granted that philosophers are uniquely qualified to handle such a project.

## 2.2 Moral Expertise

The first methodological assumption I want to address concerns moral expertise, or more accurately, the lack of it. This assumption can manifest in several forms, but at its core is the idea that the experience and study of the moral philosopher does not equip her with special insight into moral matters or give her more reliable intuitions. To be clear, this does *not* mean that moral philosophers are no better at recognizing flaws in moral arguments, or at reasoning from evidence to conclusions relating to moral matters. It is

merely a claim about the intuitions of the moral philosopher, claiming that they are not privileged over those of someone less experienced in moral questions. In this section, I examine the role of this assumption in empirical moral projects, looking at arguments and approaches that rely on this assumption. I argue that this assumption is insufficiently defended and that we have good reason to be skeptical of it. Further, without this background assumption, many empirical research projects lose much of their grounding.

One of the most common empirical approaches to moral philosophy involves expanding the number of people to whom we have access. By seeing what intuitions non-philosophers actually have with regard to philosophical problems, we test and confirm the relatively limited set of intuitions generated by professional philosophers against a wider range of people.<sup>43</sup> The motivation for this broadening range of intuitions is easy to understand, as philosophers are a fairly poor representation of the general population. Professional philosophers – a group consisting heavily (though certainly not entirely) of university professors – are a whiter and more heavily male group than the general population, with an above-average level of education, and generally falling into a particular socioeconomic class. Despite the relative homogeneity of philosophers, however, moral philosophers seek to make universal claims about how moral concepts are to be understood and employed, often through the use of intuitions. But if philosophers are such an atypical group, the concern goes, should we not be concerned that their views and intuitions are also atypical? For instance, a recent survey of (mostly American) philosophers found that nearly 73% of respondents either accept or lean towards atheism (Bourget and Chalmers 2014), compared with only 9% of the US general population (Pew Research Center 2014). So, it seems reasonable to think that philosophers have different intuitions than those around them.

It is worth noting that it is not only empirically-minded philosophers who have pointed to these sorts of problems with the ways in which traditional approaches to philosophy use intuitions. Feminist philosophers have long been critical of other

---

<sup>43</sup> For my present purposes, I am interested in experiments and surveys which consider moral intuitions, but this approach has also been applied to other areas of philosophy, as we will see in later chapters.

philosophers relying on intuitions and judgments from a highly homogeneous source: mostly affluent white men.<sup>44</sup> This objection has been at the core of various feminist accounts that have attempted to broaden the scope and sources of the intuitions philosophers use for evidence. This is what informs standpoint approaches to epistemology and philosophy of science, for instance, which “argue that what we do in our social relations both enables and limits (it does not determine) what we can know” (Harding 1995, 341). This idea is at the core of standpoint epistemology, accounts of “Strong Objectivity”, and various other feminist critiques of an overreliance of male intuition: these accounts are necessarily limited and fail to take the full range of intuitions and judgments into account.<sup>45</sup> As will be discussed further below, similar objections are raised by some feminist moral philosophers, who argue that gendered experience makes a significant difference in our experiences, and traditional accounts of ethics treat male perspectives as morally neutral or free of biases, and discard female perspectives. These feminist critiques start with the same problem the empirical criticisms do, but take their analysis and solutions in different directions. So, as my concern here is with the particular ways that empirically-minded philosophers have attempted to respond to issues with our use of intuitions, I will be focusing on their responses. Nevertheless, it is important to note that this critique is not one that is unique or new to the empirical or experimental approach, and has many roots and common influences in feminist theory.

This concern has been central to many empirically-minded philosophers, and thus has informed their approaches to ethics. Prinz is a particular proponent of looking to the intuitions of non-philosophers. He chides moral philosophers for defending a harm principle even though “ordinary people often condemn acts in which no one is harmed” (2007b, 279), and advocates “consult[ing] the intuitions of people who are not professional philosophers, such as college undergraduates” (274).<sup>46</sup> He then proceeds to

---

<sup>44</sup> Especially throughout this history of philosophy, when the discipline was even less diverse than it is now.

<sup>45</sup> See also Harding 1986, 1991, 1993; Haraway 1988, 1989.

<sup>46</sup> His choice of subject is questionable, especially given his concern that the theoretical commitments of philosophers bias their intuitions (274), and college undergraduates are hardly unexposed to philosophical

defend his account of the connection between moral judgment/sentiment and emotion by citing a study he has conducted, which he describes as follows:

I asked a group of North Carolina college students to consider two scenarios (Prinz, unpublished data). One describes a person who verbally condemns marijuana smoking but has no negative emotional response to it. The other describes a person who verbally insists that marijuana smoking is morally acceptable, but feels disgusted at those who smoke and ashamed when he himself smokes. For each scenario, subjects were asked whether the person's moral values were reflected in what they said or in how they felt. In both cases, a significant and sizable majority said that the moral value is reflected in the emotion (ninety percent in the first case and seventy percent in the second case). (Prinz 2007b, 274).<sup>47</sup>

This study – along with his earlier citation of other psychological studies which purportedly show a link between emotion and moral judgment (272-3) – makes it clear that Prinz is interested in getting at as wide a group of intuitions from as wide a range of people as he can, expanding well beyond just professional philosophers. We can safely say, then, that Prinz has some concern about the reliability of the intuitions and conclusions of moral philosophers.

Nichols takes a similar approach to Prinz, and strongly defends the idea that “the experimental investigation of folk [i.e., non-philosopher] responses is a powerful resource for charting the contours of folk concepts” (Nichols 2008, 399). One of his major projects has been to argue that the seemingly amoral nature of psychopaths challenges rationalist accounts of moral judgment (Nichols 2002). His main evidence for

---

ideas. Further, given concerns about the WEIRD nature of college undergraduates compared to the general population (Henrich, Heine, and Norenzayan 2010), they are also a problematic choice. For the purposes of this section, however, I do not wish to question the methodology of his preferred experiments. I am concerned with what his empirical approach reveals about his understanding of moral philosophy, and so we can ignore problems with the setup of his experiments/surveys and treat them as adequate.

<sup>47</sup> Again, the design of this study leaves much to be desired, especially given the practically nonexistent amount of data Prinz provides (what is the sample size? How were participants selected? Was this a large random trial, or ten students in a class he was teaching?). As before, however, I am concerned with what attempting such a study in the first place tells us about his understanding of moral philosophy, and not with any design flaws in the study itself. For my purposes here, it does not matter if the study is riddled with methodological issues or is the pinnacle of scientific rigor.



this conclusion comes in the form of empirical studies, looking at intuitions of non-philosophers to see if people generally think that psychopaths do make moral judgments despite being entirely unmotivated by those judgments (Nichols 2002, 289; 2008, 396). So, again, we see this strong desire to look beyond the intuitions of philosophers and get at what “ordinary” people think about these moral questions. This point seems obvious, but is worth bearing in mind as we proceed: there is no reason to look to intuitions of non-philosophers unless we think that philosophers and non-philosophers will vary in their intuitions.

This brings us to the important question: What does the desire to gather the moral intuitions of non-philosophers reveal about its proponents’ beliefs regarding the methodology of moral philosophy? To answer this, the first thing we should ask is why we should care if the intuitions of moral philosophers differ from those of the general population, when we may not really care about this in other scholarly contexts. If, like philosophers, physicists were considerably more likely to be white, male, highly-educated atheists than the wider population, and this caused their understanding of the concepts involved in their work to differ from the common understanding, we would presumably not think this gave us reason to doubt the results of the best work in the field. Similarly, an English professor almost certainly has a very different understanding of the works of Shakespeare than someone who has not studied them so thoroughly, but again, we would not think this gives us reason to doubt her interpretation of *Hamlet*. What, then, might explain why we might think the atypical intuitions of the physicist or literary scholar are of no particular concern, but the atypical intuitions of the philosopher are?

The most immediate answer is this: we believe the physicist when she tells us that some feature of reality is not how non-physicists understand it, or the English professor who tells us that this poem does not mean what we think it does because we think that their experience in their fields gives them a sort of expertise. The physicist better understands the laws governing the motion of bodies; the English scholar better understands how to pull meaning out of a literary work. In contrast, the thought must go, there is no such expertise possessed by the moral philosopher in virtue of her years of studying normativity. After all, if the moral philosopher has special expertise into moral

matters that allows her to recognize things that a non-expert could not, there would be no reason to consider the opinions of the non-expert, who would be much more likely to be mistaken.

Prinz and Nichols might object that their approaches do not, in fact, commit them to the idea that moral philosophers are not experts on moral matters. Certainly, neither one seems to want to directly claim that philosophers are not experts, and Nichols explicitly allows that philosophers can be – and are – experts. Rather, they both claim that the turn toward folk intuitions can help settle matters when philosophers cannot agree on a single account. According to Nichols's, "when the philosopher-experts disagree on some central feature of folk concepts, it seems particularly useful to see what the folk have to say" (2008, 399), with his particular topic of interest being the debate between motivational internalism and externalism (*ibid.*). Prinz is somewhat less charitable, but makes a similar point, claiming that, when the theoretical commitments of philosophers bias their intuitions, "[t]o escape such bias, we can try to consult the intuitions of people who are not professional philosophers" (Prinz 2007b, 274). Prinz does not explicitly grant expertise to moral philosophers, but for the moment, we can treat his account as though it does. In this case, the argument for bringing in non-expert intuitions proceeds as follows: 1) Moral philosophers have some relevant level of expertise with regard to moral matters. 2) Sometimes, these experts disagree about what conclusions should be drawn from the evidence. 3) In the event of disagreement among experts, consulting with non-experts helps to resolve the debate. Therefore, 4) when moral philosophers disagree on some central feature of a moral concept, we should turn to the folk intuitions of non-philosophers.

This argument does not work. If moral philosophers are experts at unpacking moral concepts, then bringing in non-experts to settle a disagreement is misguided at best, nonsensical at worst. It is easiest to see why this idea falls apart by considering analogous cases in which there is clear and uncontroversial expertise. Imagine a scenario in which a patient, Wade, enters a hospital complaining of symptoms which do not point to any clear diagnosis. Doctors Doom and Octopus, both highly-skilled and thorough, examine the man, performing all relevant tests. At the end of this battery of tests, two

possible diagnoses remain. Doctor Doom is convinced that Wade has toelio, while Doctor Octopus believes it to be shinitis. There are many different ways we could think the doctors should proceed from here, but one thing we would certainly not think is that the way to resolve this deadlock is to turn to people without any medical training to see what they think. Indeed, such a suggestion would seem ludicrous. Even if Wade had a team of dozens of doctors examining him who were all divided on his diagnosis, we would never think that this debate could be settled by turning to non-doctors. This is not a purely hypothetical situation, either. There is a live debate in physics over interpretations of quantum mechanics over which physicists are divided, but again, no one would think we could resolve this by turning to people with no training in physics and asking them what the correct interpretation is.

The relevant features in these cases seem to be the same as in moral philosophy: experts have conducted their work, but are divided on the proper conclusion to draw. In the case of medicine or science, looking to folk intuitions seems fruitless, yet Nichols and Prinz want to do exactly that in the moral case. What explains this difference, then? Nichols offers one possible explanation, pointing out that philosophy in general, including moral philosophy, is concerned not with the concepts of philosophers, but with folk concepts. That is to say, he rejects the idea that philosophers have or are concerned with different concepts than non-philosophers, and correctly accuses any project that claims that we are only interested in “the concepts of G.E. Moore and the like” of “sacrific[ing] any relevance outside the confines of the ivory towers” (Nichols 2008, 397). So, the argument goes, we can use folk intuitions in this case because what we are concerned with are the folk concepts of “right”, “wrong”, “good”, and such, not with special meanings of those terms of value only to philosophers. Ordinary people are able to successfully use and understand those terms, and so the intuitions they associate with them are valuable. In contrast, we can reasonably say that the physicist is employing different concepts, or using concepts in a different sense than the non-physicist. For instance, “heat” means a very different thing to the physicist than it does in ordinary conversation, and so we are not interested in what non-physicists think about that concept. So, the concepts of the expert and non-expert come apart in the physics case in a way they do not in the philosophical case. Therefore, Nichols may argue, use of folk

intuitions is acceptable to settle moral debates because non-philosophers use the same concepts we are trying to unpack, and have intuitions about them in ways they do not have intuitions about quantum mechanics.

Once again, this argument rests on a mistake. I agree with Nichols that philosophers and non-philosophers are concerned with the same concepts, and there are no special, expert concepts about which philosophers are arguing that are inaccessible to non-philosophers. What Nichols gets wrong, however, is the relationship between folk intuitions and folk concepts. We are concerned with folk concepts, but it does not follow from this that we are concerned with folk *intuitions*. Intuitions – often in conjunction with some argument for why that intuition should be trusted – serve as evidence in philosophy that help to show us what may be involved in the concept being analyzed. But the question now is, why should we consider all evidence to be equal? Remember, for the moment, we are considering the possibility that moral philosophers have some expertise in moral matters, but that consulting folk intuitions is nevertheless useful. But if philosophers have expertise and that expertise does not relate to the actual concepts being analyzed, that expertise must somehow relate to their intuitions. I discuss this further below, but if the philosopher has expert intuitions, then that must consist at least partly in being more able to see when an intuition is confused, misleading, or running two concepts together. If this is the case, however, then I return again to my original point that it is not clear what, if anything, folk intuitions are adding to the picture. And so, I must conclude, if philosophers like Prinz and Nichols want to make use of folk intuitions to establish principles in moral philosophy, it must be based on a belief that moral philosophers lack expert perspective on their subject matter.

The question we must now turn to is, should we think that moral philosophers have some expert insight into moral matters? The first thing to note is that, as Julia Driver points out, there are at least three ways in which one can be a moral expert: One can be an expert at moral judgment, more likely to arrive at true/correct judgments of what is right or wrong; an expert at moral practice, more likely to behave in accordance with the demands of morality; or an expert at moral analysis, better able to recognize what factors are morally salient, what makes something right or wrong, etc. (Driver 2013). That

philosophers have some expertise in analysis is fairly uncontroversial – it is exactly what philosophers *do*. It is also not at issue here, as the use of folk intuitions does not in any way imply that the philosopher is not better able to explain those intuitions, assess what features of them are morally relevant, and so on. Similarly, we are not – for the moment – concerned with expertise in moral practice. The underlying assumption of the empirical ethicist is not that non-philosophers are just as likely to behave morally as the philosopher; our concern is with folk intuitions as opposed to philosophers' intuitions. So, what we care about is whether or not the moral philosopher has some expertise at moral *judgment*.

In asking whether or not philosophers have expertise in moral judgment, we should consider Peter Singer's arguments in favour of moral expertise. Singer points to three sources of expertise for moral philosophers. First, expert knowledge in argumentation and detecting invalid inferences. Second, special expertise in knowing the structure of *moral* arguments, and how moral concepts are deployed. Finally, the moral philosopher has time to dedicate to moral thought that others do not (Singer 1972b, 117). These first two are forms of expertise in analysis, but the third is particularly relevant to understanding why the intuitions of moral philosophers are importantly different from folk intuitions. Consider the situations in which we would be gathering the sorts of folk intuitions to which Nichols and Prinz appeal. They are surveying intuitions in lab scenarios or via survey. So, the conditions do not allow for reflection or analysis, but are rather quick, initial intuitions. But this is not what is going on with philosophical intuitions. The intuitions of the moral philosopher – or at least those that make it into print – are not immediate, snap judgments; they are not unreflective intuitions. Rather, the intuitions used as evidence by philosophers are ones which we have time to consider, examine for morally irrelevant or inconsistent factors, and endorse on reflection, or reject the intuition if it is problematic. On this understanding, the intuitions of the moral philosopher are a factor not just of her moral judgment, but of her moral analysis as well, and as we have already discussed, moral philosophers have some expertise in moral reasoning. Are the unreflective intuitions of non-philosophers just as reliable as those of philosophers? Possibly, but this is not what is actually at issue. Rather, the question is, do we have reason to think that the considered intuitions of moral philosophers are more

reliable than the unreflective intuitions of non-philosophers? In this case, given the expertise moral philosophers have in moral analysis and the ability of philosophers to apply their expert analysis, thereby refining their intuitions, we must conclude that – absent some defeater – the intuitions of moral philosophers are more likely to be accurate, and thus that philosophers do have relevant moral expertise.

Eric Schwitzgebel and Joshua Rust have run a series of studies that appear to challenge the notion that moral philosophers have any sort of expertise. They have gathered self-reported moral attitudes and behaviours of both moral and non-moral philosophers and professors in non-philosophy departments, looking to see if moral philosophers were any more likely than others to engage in the behaviours they claimed were morally required (or prohibited). Across the board, they found that moral philosophers were no more likely to actually engage in the behaviours they claimed were morally obligatory than were the other groups (Schwitzgebel and Rust 2014). Given this information, we may legitimately ask about the expertise of moral philosophers. If the moral philosopher does not behave any more morally than anyone else, are they truly experts?

We may immediately respond to this criticism by pointing out that this merely shows that philosophers are not expert practitioners, as mentioned above, and does not challenge their expertise in moral judgment or analysis, which are the only things required to undermine the reliance on folk intuitions. This argument, however, seems too quick. They may be three areas of expertise, but to show that a lack of expertise in moral practice is irrelevant, we must establish that such a lack is not evidence of a lack of judgmental or analytic skill. After all, it certainly seems reasonable to expect someone with expert knowledge in some field to also be able to apply that knowledge in their practice.

The main reason we might be concerned about moral philosophers failing to live up their self-ascribed obligations stems from the fact that moral beliefs/knowledge are supposed to be motivating. If I believe that I have a moral obligation to give a portion of my income to charity, or that eating meat is morally unacceptable, then I ought to be

motivated to adjust my behaviour accordingly. If I don't donate to charity or continue to eat meat, one might well think that I don't view my behaviours as morally unacceptable, and thus question if I am actually committed to the norms I claim to endorse. So, if moral philosophers do not behave any better than anyone else, we may doubt that they have actually been any better in their moral judgments than anyone else, and thus do not have any real moral expertise. The question before us, then, is this: Does a moral philosopher's failure to behave in accord with the demands of morality in fact indicate a lack of moral judgment or analytical skill? If we want to answer in the negative – as I do – then we need some sort of explanation for how one could make the appropriate moral judgments, be motivated to act in accord with those judgments, yet nevertheless fail to do so. Fortunately, such an explanation does exist.

It is not particularly difficult to explain how knowing what is morally required and consistently acting in accord with that requirement can come apart.<sup>48</sup> Moral philosophers since Plato have discussed the impact of weakness of the will and any number of other factors that may interfere with an agent acting in accord with morality's demands. In such cases, the agent knows perfectly well what morality demands of her, but nevertheless fails to act on it. These are cases in which an agent has some level of motivation to act in accord with what they judge is best, but that motivation is overridden by other factors. We can easily explain the apparent failure of moral philosophers to act morally in these terms. One may, for instance, judge that it is good to call or visit one's mother frequently,<sup>49</sup> but be overcome by inertia when the time comes for a visit, and opt not to make the journey. Exactly why weakness of the will occurs, and exactly how it might be that someone could form the judgment that action A is better than action B, but nevertheless go on to perform B is a heavily-debated question, and is certainly not a problem that can be solved here. For our purposes, what matters is that weakness of the

---

<sup>48</sup> As we saw from Schwitzgebel and Rust (2014), discussed above, these do in fact frequently come apart, and ethicists frequently do fail to behave any more morally than anyone else, despite any expertise they may have in telling people what is morally right.

<sup>49</sup> An example from Schwitzgebel and Rust 2014, 303

will is something that occurs, and the failures of moral philosophers to behave morally can be plausibly explained as instances of it.

If this is true, then we can ask a slightly different question: should moral philosophers – if they are legitimately experts – be better at overcoming a weak will than other philosophers, or the general public? In this case, I don't see any reason to think that they should be. Moral philosophers tend to be more interested in practical reason than other philosophers, and so may have the terminology to describe the phenomenon, but actually dealing with it is a question of changing unhealthy patterns of thought and behaviour. This is a task that moral philosophers are not experts in, nor should anyone expect them to be; fixing unhealthy patterns of behaviour is a problem for a psychologist, not an ethicist.<sup>50</sup> We should not expect moral philosophers to be experts at dealing with weakness of the will, and so we should not be surprised if they are prone to be weak-willed and fail to act in accord with the demands of morality just as much as anyone else. That moral philosophers are no more likely to behave morally, then, is not evidence of a lack of expertise in moral matters. Rather, it indicates only that moral philosophers are just as likely to be weak-willed as everyone else, which should not be surprising. If an ethicist fails to perform some behaviour that she says is morally required, our response should not be to question whether or not she properly understood what was morally required.<sup>51</sup> Even if moral philosophers are just as likely to fail to live up to what is morally required as other people, this need not shake our confidence in the judgmental or analytical expertise of moral philosophers. And so we can conclude that moral philosophers have expertise which sets their intuitions apart from the intuitions of non-philosophers.

Once we have established that philosophers are moral experts, a serious problem arises for the empirical project. The use of non-philosophers and folk intuitions to

---

<sup>50</sup> Analogously, a physicist could tell you exactly what happens when a human body is exposed to ionizing radiation, and why that radiation damages living tissue, but if you want your radiation burns treated, you should see a doctor. This is not evidence that the physicist is not an expert on radiation.

<sup>51</sup> Absent some factor that indicates she failed to grasp the moral requirement.



challenge or undermine traditional work in philosophy cannot be justified without the background assumption that folk intuitions are on a par with those of moral philosophers. As we have seen, however, this is not the case. The special experience and knowledge of moral argumentation and analysis of the philosopher reinforces her intuitive judgments, making them more reliable than folk intuitions. Given this, we should reject the assumption underlying the empirical project, and thus also be highly skeptical of any data arising from surveying folk intuitions.

## 2.3 Describing or Prescribing Norms?

Ultimately, the question of whether or not philosophers specializing in ethics have any special expertise in answering moral question is an epistemological one. At issue is what counts as evidence for and against moral claims, and how strongly they count. A more fundamental question about the methods and aims of moral philosophy arises when we ask if it should be considered a descriptive project, or a prescriptive one. That is to say, are we trying to give a descriptive account, matching our account of ethics with the norms and values people actually hold? Or is the aim of moral philosophy to argue for an ideal (or at least better) set of norms we *ought* to hold? The answer need not be exclusively one or the other. Indeed, a full account of ethics will almost certainly have both descriptive and normative elements. The question, rather, is what is our *primary* aim? Those taking the empirical approach see it as descriptive,<sup>52</sup> but the history and current practice of moral philosophy is more complicated than this. In this section, I will argue that, while there is a significant descriptive component, the major project of moral philosophy is a normative one, and that this presents a significant challenge to the empirical approach.

To begin, I should clarify what I mean by a descriptive account of ethics. By a descriptive account, I mean one that holds the following claims:

---

<sup>52</sup> I do not claim that these philosophers explicitly claim that their account is a descriptive one, nor even that they necessarily *think* they are giving a descriptive account. Prinz, for example, explicitly argues against the “unifying essence” approach to moral philosophy (2007b, 279). Rather, I claim that their approach to ethics necessarily commits them to a descriptivist account, even if they claim to reject it, as I have argued above.

- 1) The goal of moral philosophy is to describe the norms that people actually hold.
- 2) The norms that actually exist can be determined by analyzing the ways in which people talk/think about morality, and the proper linguistic use of moral terms.
- 3) Any supposed norm or account of morality that does not adhere to what “we” actually think is mistaken.

The prescriptive account, on the other hand, holds that moral philosophy is itself a normative enterprise. The concern is not with generating a list of norms people actually hold, or to find a unifying essence that brings our moral judgments together. Rather, we are searching for the norms that people *should* hold, the following of which would lead to a “better” world.<sup>53</sup> On this understanding, ethics is an ameliorative project, arguing for ways that our system of norms can be improved.<sup>54</sup>

### 2.3.1 The Descriptive Account

The descriptive understanding of morality is common among empirically-minded philosophers. Prinz, for example, derives his motivation for an empirical inquiry into moral matters from such a descriptive account, stating,

If the good is that which we regard as good, then we can figure out what our obligations are by figuring what our moral beliefs commit us to. Figuring out what we believe about morality is a descriptive task *par excellence*, and one that can be fruitfully pursued empirically. Thus, normative ethics can be approached as a social science. (Prinz 2007a, 1, original emphasis)

---

<sup>53</sup> “Better” here should be understood as meaning “more in line with an ideal morality”, and not necessarily “better” in the consequentialist sense of producing the more desirable outcomes for a larger group of people (although these two senses will be the same for the consequentialist).

<sup>54</sup> It is important to note that the descriptivist view of ethics does not necessarily commit its holder to an empirical or neuroethical approach. While this account is certainly compatible with an attempt to discover norms through empirical inquiry, a descriptivist could also take a more traditional approach, such as by attempting to discover the actual norms through conceptual analysis. As we will see below, a number of historical philosophers can be read as trying to give exactly such a descriptivist account.

Given his metaethical commitment to an Agent Relativist account,<sup>55</sup> Prinz adopts this descriptivist take on morality, which motivates and justifies his use of empirical data in investigating moral questions. Indeed, understanding morality as a prescriptive enterprise could not be an option for Prinz, since on such an account, the good is not – or at least is not necessarily – what we think it is.

We also see Prinz's fondness for a descriptive understanding of ethics in his use of empirical studies as support for his positions. One of the clearest examples comes from his "Normativity Challenge" to virtue ethics. Prinz is concerned with the question of where virtue ethics derives its normative force and attacks it by pointing to cross-cultural studies which indicate that there is no broadly accepted set of virtues or factors which contribute to well-being or flourishing (Prinz 2009, 134). This, he argues, undercuts the idea that there is one universal set of virtues (138). This sharply contrasts the attitudes taken between a descriptive and prescriptive approach to morality, as there are two ways to respond to such data. A prescriptivist has no particular trouble with disagreement. On the prescriptivist account, we can expect that some of the norms to which people adhere – even ones which are supported at the level of their culture – will turn out to be wrong. There may be widespread disagreement about what virtues best contribute to flourishing, but which of these should be adopted can be determined through typical philosophical argumentation. For the descriptivist, the only choice is to say that the existence of these competing virtues, with no agreement as to which ones are the best or universal ones means that we must discard the idea of a single account of virtues. We are attempting to describe the set of virtues that are actually held as universally good and contributing to flourishing, but no such set exists, so we do away with the idea. It is the latter of these two responses which is adopted by Prinz, and so we can safely say that he is defending a descriptivist understanding of moral philosophy.

Churchland takes a similarly descriptive approach, but for a different reason. For Churchland, the purpose of an empirical investigation is to connect moral philosophy to

---

<sup>55</sup> i.e. What is right or good is determined by what is thought to be right or good by the person making the moral claim and the relevant social or cultural groups (ibid.).

information about evolution, the brain, and neuroscience, and prevent it from becoming a matter of, “mere, albeit confident, opinion” (2), which she sees any moral project not tethered to the “hard and fast” (ibid.) – by which she means evolution and science – as being at risk of becoming nothing but such opinion. Her project, then, is to make use of such data to “meaningfully approach the question of where values come from” (3), but not with answering particular moral questions about what we ought or ought not to do. She is committed, then, to investigating the actual, empirically-confirmable source of our moral thinking, but this does not actually tell us whether not she is committed to a descriptivist understanding of ethics. What will determine this is what Churchland sees as being the value of such a project.

There are, broadly speaking, two major reasons we may want to determine the evolutionary or neuroscientific bases of moral judgments, which can be understood as ways of attempting the second and third possible sociobiological endeavours identified by Kitcher (Kitcher 2006, 576). In the first case, we may want to know the origin of our various moral intuitions so that we can know which ones come from reliable sources, and which we should be regarding with a greater air of suspicion, potentially revising or discarding. In Kitcher’s words, these facts, “in conjunction with moral principles that we already accept, can be used to derive normative principles that we had not yet appreciated” (ibid.). Flanagan, Sarkissian, and Wong, for example, argue for such a position when defending a naturalized ethics which can “explain *why* some norms (including norms governing choosing norms), values, and virtues are good or better than others” (Flanagan, Sarkissian, and Wong, 16, original emphasis). They argue that empirical knowledge can allow us to favour a set of norms that “is suited to modify, suppress, transform, or amplify some characteristic or capacity belonging to our nature – either our animal nature or our nature as socially situated beings”, tracking norms which “serve to regulate and promote human social life” (ibid.). Here, the already accepted moral principle is the idea that what is good is what regulates and promotes human social life, and new data about the evolutionary basis of our norms tell us (or help tell us) exactly what fits that description, and what fails to. Such a project would fall in line with the prescriptive account of ethics I have sketched. Our goal is to find out not what it is we already believe or are committed to, but rather what norms we should commit to.

Depending on one's background moral principles, how strongly one wishes to tie morality to the empirical data, and other metaethical commitments, there are a variety of ways to separate the reliable intuitions from the unreliable. If we think morality strongly tracks (but does not necessarily perfectly match) our evolved nature and tendencies, then we may conclude that any norms not easily explicable in terms of adaptive traits may be reasonably regarded with skepticism. If, on the other hand, we think of our evolutionary history as something which may lead us astray from our proper, rational judgments and perceptions of what is morally right, then we may think that judgments that stem from attitudes which were useful in our evolutionary history, but less relevant to modern human life are to be treated with skepticism. For example, we may conclude that ingroup/outgroup distinctions – despite informing some of our moral judgments – are an outdated evolutionary mechanism stemming from times when caring about the well-being of one's own tribe was much more important than it is now. From this, we may further conclude that the moral sentiments which stem from this evolutionary trigger, such as the view that we are obligated to put the needs of people close to us above those of people with comparable – or even greater – need in distant parts of the world, are one that we are better off without.

Alternately, we may want to find the basis of moral judgments because we believe that doing so will determine what the “correct” normative account is, and “explain what ethics is all about” (Kitcher 2006, 576).<sup>56</sup> For instance, if we were to discover that moral judgments stem from judgments about some relatively narrow conception of fairness, and we are committed to a descriptive account of morality, we may conclude that this particular concept of fairness is all there is to ethics. This seems to be the approach taken by Churchland. She is concerned less with the possibility that understanding the sources of our moral intuitions may show us which ones are to be trusted or treated with caution, and more with “how to tether ideas about ‘our nature’ to the hard and fast.” (Churchland, 3). As part of her argument against the is/ought gap, she hypothesizes that naturalism will

---

<sup>56</sup> Kitcher presents an argument for why such an endeavour is misguided (579-81), but for our present purposes, we can leave this concern aside and grant that, if we're seeking merely to describe actually existing/accepted norms, discovering facts about human beings can at least potentially do this

“find the roots of morality in how we are, what we care about, and what matters to us” (6). This already seems to suggest that she is interested in tying morality to what we actually care about, rather than what we should, but we can perhaps see her commitment to a descriptivist understanding of morality most clearly in her criticism of what she calls “maximizing consequentialist” views, and particularly those of Peter Singer. She suggests that we can question whether or not, “Socrates or Confucius, both examples of morally good humans, calculate[d] as maximizing consequentialists decree” (178), and accuses the consequentialist of holding a view which can run afoul of commonly held and accepted moral beliefs (179). This seems to suggest that she is thinking of morality in descriptivist terms. Singer’s account is wrong because it does not match up with the norms that people actually hold, and may not describe the moral practices of individuals that we can agree were of good moral character. If this is the reason she rejects his account, then she must be committed to some form of the idea that what is morally correct is what people or society generally accept as being morally correct, which is the core of the descriptivist picture.

Prinz and Churchland are two particular examples here, but they are certainly not the only ones sketching a descriptive account of ethics. Any account that is deeply concerned with determining what people actually think, or what norms they actually adhere to is going to have some strongly descriptive element. This is true whether we are conducting cross-cultural surveys to see how different groups respond to moral dilemmas looking for points of widespread agreement, reducing morality to some sort of neurochemical reaction in the brain, or just arguing that the conclusion to some moral question presented by a fellow philosopher is mistaken because her answer does not coincide with what people would generally find morally acceptable. As a result of this, any account of morality which holds that we need to test and get clear on what people actually think, as many of the empirical accounts do, is going to be committed to the idea that the task of moral philosophy is, on some level, descriptive. But is this the case? Is moral philosophy best understood as a descriptive undertaking?

### 2.3.2 The Moral Project

If we want to determine how to best understand the project of moral philosophy, we will be best served by looking to its practitioners. By considering paradigm cases of moral philosophy and examining the approach they take, we can begin to see the understanding these philosophers have of the practice of moral philosophy. The accounts we consider must have two important characteristics. First, the accounts with which we are concerned must be fairly “typical” projects, which is to say, ones that are engaged with by moral philosophers, treated as generally taking part in the moral philosophical tradition, and so on. Second, while we are not concerned with whether or not they are correct, the concerns cannot be that they fundamentally misunderstand what moral philosophy is, or how it is done. So, for example, while we may doubt that Aristotelean virtue ethics is an accurate account of our moral duties, we do not doubt that he is concerned with the same broad understanding of moral philosophy as other ethicists. With this in mind, there are four particular examples of moral projects I wish to consider: Kant’s description of his project in *Grounding for the Metaphysics of Morals* (*MM*), Henry Sidgwick’s argument against common sense morality from *The Methods of Ethics* (*ME*), Peter Singer’s work, with particular attention to “Famine, Affluence and Morality”, and finally, feminist work on care ethics. I argue that all four of these are explicitly prescriptive projects and – more importantly – we can take general lessons from these examples to argue for a prescriptive understanding of moral philosophy.

Kant opens the *Grounding* by laying out the aims of his project, and distinguishes the moral laws with which he is concerned from the practical rules derived from experience (*MM*, 389). Given some of his remarks on the role of his project, we could be forgiven for thinking that Kant is attempting a descriptive undertaking. He claims that the prohibition against making a false promise is “within the moral cognition of ordinary human reason” which will “always have it actually in view and does use it as a standard of judgment” (403-4). Further, he claims that these principles are clear so long as we do not “try to teach reason anything new but only make it attend [. . .] to its own principle” (404). This makes it sound as though Kant sees his project fitting into the descriptive approach I described earlier. Looking more carefully, however, this reading begins to

collapse. At the outset, Kant makes it clear that he is concerned with “the idea and principles of a possible pure will and not the actions and conditions of human volition as such” (390-1). In other words, Kant’s account is in accord with what we *would* think if we were all being purely reasonable about morality (which Kant thinks we should), but he does not think this is how we *actually* think about it. On this reading, Kant is not attempting to put forward a description of actually held norms, but the norms we should hold, and thus is engaged in a prescriptive project.

Looking to the utilitarians, we should expect to find them attempting to argue for the norms we should hold over the ones we may in fact hold, given the political history of the position. And indeed, this is exactly what we find. Sidgwick’s arguments against common sense morality (CSM) make it clear that he is attempting to move beyond a description of how we think about morality. He lays out four conditions that an account of morality must meet (*ME*, Bk. III, ch. XI, §2), then argues that the intuitions of CSM fail to live up to these criteria (*ibid.*, §3-9). Sidgwick is not trying to find something that can unify or ground ordinary accounts of morality. Rather, there is some desiderata that an account needs to be satisfactory, and if our ordinary descriptions do not fit this, then they need to be rejected or revised. If this is the case, though, then Sidgwick is also undertaking a clearly prescriptive project.

One of the clearest and most explicit examples of a moral philosopher tying their project not to what norms we do hold, but to what we should hold is Singer. Certainly, his understanding of morality is based on *some* level on norms we do in fact accept. The claim that, “if it is in our power to prevent something bad from happening, without thereby sacrificing anything of comparable moral importance, we ought, morally, to do it” (Singer 1972a, 231) is intended as an uncontroversial, and broadly accepted one (*ibid.*). Similarly, his famous example of the drowning child in the shallow pond is meant to match uncontroversially with ordinary, broadly-held intuitions (*ibid.*). Singer is quite explicit, however, that his thesis is that, “the way people in relatively affluent countries react to a situation like that in Bengal [during the Bangladesh Liberation War of 1971] cannot be justified; indeed, the whole way we look at moral issues – our moral conceptual scheme – needs to be altered” (*ibid.*, 230). This is a theme that has continued



throughout Singer's work: he is concerned not with what our morals are, but what they should be.

Finally, care ethics, which has been defended by some feminist ethicists,<sup>57</sup> is very much founded on a rejection of what moral views people actually hold, arguing instead for how we ought to think or morality. Carol Gilligan's (1982) account of care ethics is explicitly a response to Freudian (and especially Lawrence Kohlberg's) accounts of morality and moral development that privileged "male approaches" to development over "female" ones.<sup>58</sup> In virtue of this origin alone, we can already see that this takes a fundamentally normative approach. It is, from the outset, pushing back against established accounts, and trying to establish an alternate way of looking at ethics. It is not, in other words, just a description of the moral positions people actually hold. Care ethics is sometimes framed as better representing or capturing how women think about or engage with moral duties, but even then, the project should not be thought of as a descriptive one. Even though care ethics attempts to give better representation to "female" methods of moral reasoning, characterizing it as merely a description of how its defenders characterize women's thoughts about ethics is misleading at best. Care ethics starts from a purportedly "female" perspective, to be sure – as do the standpoint theories briefly discussed above – but the claims about the sources and nature of our moral duties are meant to be every bit as broadly and universally applicable as those of any other moral theory. And so, once again, we have a prominent account of moral philosophy which is best understood as trying to prescribe a set of norms we ought to hold, and not merely accurately report on the norms we really do hold.

These are, of course, just four examples, but we can extract some general principles from them. The first conclusion we can draw from these is that – at the very least – several major traditions in the history of moral philosophy adopt a prescriptive

---

<sup>57</sup> And strongly criticized by others

<sup>58</sup> As is the case for standpoint approaches to epistemology and philosophy of science discussed above, this assumes (somewhat controversially) that gendered experiences make a significant difference, and thus that we can meaningfully distinguish between "male" and "female" approaches to reason.

approach. Already, this should be enough to make us suspect that philosophers like Prinz and Churchland are engaged in a different kind of project than other moral philosophers. Even more importantly, while these projects may be considered particularly ambitious or broad in scope, none of them would be considered outlandish or atypical. That is to say, they are all fairly normal (in a sense) projects in moral philosophy. This indicates that taking a prescriptive approach does not mark a philosopher's work as being particularly odd or unusual, which – while certainly not conclusive – at least suggests that such an approach is not an uncommon one.

Étienne Brown, responding to Prinz's empirical criticisms of virtue ethics, suggests the following principle, which he takes to be generally acceptable: "It may be morally right for a specific agent to  $\phi$  even if empirical cross-cultural comparisons show that there is substantial disagreement about whether she ought to  $\phi$ " (É. Brown, 145). As we have seen, Kant, Sidgwick, and Singer would all be willing to accept this view. What's more, this would seem to be an acceptable premise to most moral philosophers. We can see this if we consider a stronger version of the claim:

It may be morally wrong for an agent to  $\phi$  even if there is broad cultural/cross-cultural agreement that she ought to  $\phi$ .

While this may seem a bold claim, it follows from several cases which would seem to be broadly acceptable. Singer's argument that it is wrong for well-off individuals to fail to give substantial amounts to charity in the third world – regardless of what our ordinary intuitions hold – is an instance of this principle. Similarly, consider the claim that slavery is wrong. This (I hope) would be taken as acceptable by most moral philosophers. But what's more – although not uncontroversial – many would also accept the stronger claim that slavery was wrong even when it was broadly practiced. So, it was wrong in ancient Greece, wrong in pre-Civil War America, and would be wrong if any future society which judged it acceptable. If this claim is acceptable, though, then we are committed to some form of the claim that an action can be wrong even if people do not judge it to be wrong. Such a principle is the core of a prescriptivist approach to morality, and antithetical to a descriptivist approach. So, if this principle would be generally accepted,

we should conclude that the dominant understanding of moral philosophy is prescriptive and not, as Prinz and Churchland understand it, descriptive.

There is one more general question we may want to ask when considering how we ought to understand moral philosophy. If moral philosophers believe the aim of moral philosophy to be simply laying out the views people actually hold, why would this project be worthwhile? In most ordinary cases – that is, situations we encounter day-to-day, rather than the odd moral dilemmas of philosophical thought experiments – people are fairly good at knowing what they think they should do, and particularly good at knowing what is and is not within the norms laid out by their society. The difficulty lies in trying to figure out what overarching moral code may exist, and what views need to be revised or rejected in light of that code. If the goal is just to explain what people actually think, then moral philosophy would seem to contribute nothing beyond what non-philosophers can work out on their own. But presumably the question for moral philosophers is not, “what do we think is good?”, or, “what do we think is right?”, but rather, “what is good (or right)?” or, “what should we think is good (or right)?” If this is the central question of moral philosophy, then we are firmly committed to a prescriptive understanding.

By committing a descriptive understanding of the moral philosophy, empirical projects like those of Prinz and Churchland are ultimately answer a completely different question that traditional ethicists. They could attempt to claim that, despite only presenting evidence regarding norms people actually hold, they are nevertheless attempting to answer the question of what actually is morally good, but such an argument will not work. To move from a description of norms people already accept to any sort of claim about what we ought to accept is simply to offer the sort of is/ought bridging argument I discuss in the next chapter. As I will show there, arguments of that type are not viable. The only remaining option for the empirical ethicist is to give some sort of argument for why it is we should believe that what is good maps onto what people (or a majority of people) actually think is good. But this, it must be noted, would and could not be an empirical project without being question-begging. Instead, the only way to argue that the correct set of norms is identical to whatever people actually hold would be to lay

out a metaethical thesis defending that position. Such a project is certainly possible, but it is also a very traditional project of moral philosophy, and not one which is helped in any substantial way by the sort of empirical work that philosophers like Prinz and Churchland present. So, we must conclude, the assumption of the empirical ethicist that moral philosophy is an ultimately descriptive project that can be resolved strictly through the examination of norms people actually adhere to is a misguided one, fundamentally incapable of answering the sorts of questions moral philosophers are asking.

## 2.4 Moral Debunking

The final empirical assumption concerns not a positive attempt at constructing a moral theory, but a way of making a negative argument against some theory. Specifically, we are concerned with attempts to falsify moral theories by empirically investigating their practical consequences, and showing that those consequences do not obtain in the world. In the abstract, these debunking arguments generally take the following form:

P1) If moral theory  $x$  is true, practical consequence  $y$  will follow.

P2) Empirical testing shows  $\sim y$  (or does not support a finding of  $y$ ).

Therefore, C)  $x$  is false.

As an outlandish example, consider the descriptive moral theory which holds that the only thing people actually value is talking about, thinking about, or living in Utah. Call this theory, “Utahlitarianism.” A group of Salt Lake City psychologists and sociologists, inexplicably intrigued by this theory, set out to see if it is true. They test people to see what they value, and find that at least some people also value wellbeing, close relationships, family and so on. They conclude that Utahlitarianism, promising as it seemed to be, must be false. At first glance, this appears to be a more promising approach than the others, as this method of argument would seem to follow from the generally accepted “ought” implies “can” principle. If we can prove that something which a moral theory claims ought to be the case is not and cannot actually be the case, then we can conclude that it is inaccurate to say that it ought to be.

In this section, I explore attempts at debunking moral theories and challenge the assumption that informs P1. Namely, that philosophically interesting moral theories have these empirically falsifiable practical consequences in the first place. I consider two theories that attempt to debunk moral theories: Joshua Greene's attempt to empirically refute Kant's rationalism, and various recent empirical attempts to refute the principle that "ought" implies "can." I argue that neither of these empirical cases could be considered to have successfully refuted their targets, as their results can all be adopted and explained by proponents of the views they criticize without any significant alterations to their theories. I further argue that, given the failures of these empirical projects along with what we have seen in the previous two sections, we can conclude that no philosophically interesting prescriptive moral theory can be refuted by falsifying its empirical consequences.

In "The Secret Joke of Kant's Soul", Greene presents a series of experiments that show a connection between emotional responses and moral judgments typical of deontology. One of his major examples involves comparing responses to two variants of the trolley problem. In studies, people tend to make the consequentialist judgment in the ordinary trolley case (i.e., you ought to flip the switch to save the 5, killing one in the process), but make the deontological judgment when it comes to pushing someone off a footbridge to stop the trolley (Greene 2008, 42). His results show that people who make the typically deontological judgment show high activation in regions of the brain associated with emotion, while the consequentialist judgment was connected to activation of "cognitively" associated reasons (ibid., 43-4). He argues that this shows deontology to be a sort of "moral confabulation" (ibid. 63) which, rather than providing a rational basis for morality, as it claims to, offers up post-hoc rationalizations for automatic emotional triggers (ibid.). This, he concludes, "makes trouble for people in search of rationalist theories that can explain and justify their emotionally driven deontological moral intuitions" (ibid., 75).

Setting the actual experimental data aside, Greene's argument is fairly straightforward. Deontological accounts of morality are presented as being based on pure reason/logic/etc., and importantly are not supposed to stem from emotional responses or

concerns for consequences. When we look at what is actually going on in the brain when someone makes a deontological judgment, however, the emotional parts of their brain seem to be at work. Deontology, therefore, rather than being based on reason, is merely a rationalization of our emotional instincts. So, we should be skeptical of deontological accounts of morality as they evidently misunderstand how people make moral decisions.

While this disconnect between the apparent motivations of those who make typically deontological judgments and what deontologists take to be the source of normativity is interesting, it is not a devastating objection to the deontologist. The first thing to note is that “deontological” is functionally defined by Greene as “judgments in favour of characteristically deontological conclusions” (ibid., 39). In this case, by definition he is judging only the conclusion and the outcome. But for many deontologists – and certainly those most directly influenced by Kant – looking to the outcome is precisely the wrong thing to do. We can tell if someone is living up to the demands consequentialism places on them by looking at the outcomes of their actions, but the same is not necessarily the case for deontology. For Kant and many deontologists, it is motivation, and not particular actions, that is relevant. In this case, for the examples Greene cites, the motivation would be something like the satisfaction of certain emotional states. But in this case, as this is not a suitable deontological motive, the test subjects have only made a judgment that coincidentally happens to match up with what a deontologist would do, and have not made a properly deontological judgment.

Greene anticipates such a response, and here his point about deontology being a sort of rationalization comes in. He argues that you need to find two factors to show that someone is rationalizing judgments through something other than the factors they claim. First, find what predicts their judgments. Second, show that this factor is not plausibly related to what they claim is the basis of their judgments (ibid., 67-8). This is what he believes is going on in the case of Kantian morality. What Greene appears to miss, though, is that this kind of result would not surprise Kant in the least. Kant was well aware of the demandingness of his moral system, and the difficulty in performing an act not out of inclination or habit, but purely out of duty. In fact, he takes the problem slightly further than Greene does. Kant admits the possibility that “there cannot be cited a

single certain example of the disposition to act from pure duty” (*MM*, 406), and admits that there is no possible way to know for sure that a given action was actually done from moral duty (407). Despite this, he maintains that even if there had *never* been a moral act – not just that we were not sure one had ever occurred, but there has certainly never been a single one – reason would still compel the understanding of morality he lays out. For Kant, then, Greene’s empirical results are perfectly in line with what he already believes. Being morally good is difficult, and many people will fail to live up to it, so of course the people whose moral judgments we study are likely to be acting on the wrong sorts of motivations.

Greene’s mistake lies in taking Kant’s moral philosophy to be a descriptive account. If Kant believed he was describing the ways people actually reason and the factors that actually influence their judgments, then the empirical data Greene cites would be a problem. But as was discussed in the previous section, this is not how we should understand Kant, and so Greene’s objections miss their mark. We may believe that a moral theory which holds that there may never have been a single properly moral act is a problematic one, or think that it’s an issue that it is potentially impossible to tell if an act is morally good. Indeed, these both seem like potential issues. But criticisms of this type are philosophical arguments concerned with the desiderata of moral theories and what makes a given theory correct, and not empirical arguments. Empirical data may be informative, but the argument that a good moral theory ought to have characteristics x, y, and z, which Kant’s theory lacks is one that must proceed through ordinary philosophical means.

Another target of moral debunking is the “ought” implies “can” (OIC) principle; the idea that our moral obligations are restricted by what is within our ability. Two recent articles have conducted experiments which challenge this notion. Buckwalter and Turri (2015) present a series of experiments showing that people still ascribe obligation to an agent even in cases where they are unable to perform the obligatory action. Chituc et. al (2016) follow this up with further experiments, showing that “ought” judgments track judgments of blameworthiness, not ability. In both cases, the experimenters conclude that OIC is false. Chituc et. al argue that, as OIC is supposed to be an analytic principle, it

should “follow necessarily from the concepts expressed by the words ‘ought’ and ‘can’” (Chituc et. al, 23), in the same way that “unmarried man” follows necessarily from “bachelor.” According to Chituc et. al, “no one with the relevant concepts of ‘ought’ and ‘can’ should [accept that someone is unable to perform an action, but still insist that they ought to] if ‘ought’ analytically or conceptually implies ‘can’” (21). So, we are to conclude, since people who understand and can competently employ the words “ought” and “can” employ the terms in ways that are incompatible with OIC, “can” does not follow analytically from “ought”, and so, OIC is mistaken.

There are two problems with this approach. First, and most importantly, it rests on a highly questionable understanding of analytic truths. The argument claims that, if OIC is analytic, then people should not respond in a way that indicates that the two terms are separable, in the same way we would not expect people to respond that Charlie is a bachelor, but deny that he is an unmarried man. The mistake lies in thinking that all analytic judgments must be transparent, and thus impossible for competent speakers to consistently misapply. This is true of many of the paradigmatic analytic statements we use as examples, but it is a significant overgeneralization to think that *all* analytic statements must be similarly transparent or obvious. The truths of mathematics and logic are analytic, but not transparent. People who understand how to use mathematical or logical terms often get answers wrong, or do not know if a given statement is true. The average person, shown a statement of the incompleteness theorem, would in all likelihood say that they do not know whether or not it is true (unless they have some training in math), but the conclusion that follows from this would surely not be that mathematics is not analytic. Following this line, we could argue that OIC is indeed analytic, but like many other analytic truths, it is not a transparent one. And so, we may perfectly reasonably expect ordinary people without some experience in the philosophical analysis of the concepts to make mistakes in applying them, and OIC is not debunked by people making mistaken judgments about uses of “ought” claims.

The second problem relates to a matter already discussed. The experiments showing that common sense intuitions do not support OIC rest on the assumption that philosophers are not experts. Moral philosophers accept OIC because of arguments in its



favour, and philosophical intuitions supporting it. Objecting that test subjects have intuitions which conflict with OIC is only relevant if the intuitions of non-philosophers are as good as those of philosophers. If they are not, then there is no reason to examine folk intuitions. As I have already argued above, however, this assumption is unwarranted. And so, as above, we should not consider the experimental data to have debunked OIC.

In the two examples we have seen, empirical data did not successfully refute an existing moral theory, but can we generalize from this? After all, the “Uthlitarianism” example seems to show that at least some moral theories can be debunked by appropriate empirical evidence. If this is the case, then it would seem that we cannot say that the overall assumption that the practical consequences of moral theories can be empirically debunked is mistaken. Some experiments, such as the two I have discussed, may not successfully debunk their targets, but others might.

To see why this is not the case, and why attempts at moral debunking are ultimately misguided, we need to note several things. First, Uthlitarianism is a bizarre system of morality, and can and would be rejected on philosophical grounds long before we decide to do any empirical investigation into its consequences. More importantly, Uthlitarianism is a descriptive project, and as I argued above, moral philosophy is best understood as a prescriptive enterprise. But a prescriptive philosophy, which is concerned with what people should do, rather than what they will do, simply does not have the same sort of empirical consequences that a descriptive one does. Descriptive, scientific theories, even very simple ones, have empirical consequences which we can test through observation. A consequence of Galileo’s account of gravity is that a bowling ball and a baseball, dropped from the same height, ought to hit the ground at the same time. We can look at the world, conduct experiments, and if this turns out to be untrue, reject the theory. But what are the practical, empirically observable implications of, for example, Aristotle’s account of virtue? Certainly, it has implications for what we *should* do (cultivate virtue, seek a mean between extremes, and become more virtuous by behaving as a virtuous person would), but it is not clear that there are any implications regarding what *is* the case. We may question whether or not practicing virtuous actions makes you more likely to habitually engage in them, but this would not challenge the idea that being

morally good requires cultivating virtues. It only challenges the account of how we cultivate them. Similarly, a utilitarian may have many ideas about what sorts of policies, systems of government, and so on will result in producing the most happiness. We may later empirically discover that those particular things are not terribly conducive to happiness, and other policies would work better, but it would be a mistake to think this empirical information debunks *utilitarianism*. Now, were the utilitarian to take a descriptive approach and claim that *in fact* people make all of their moral decision in light of what they believe will produce the greatest overall happiness, we could empirically test and debunk that theory. Much like Utilitarianism, however, that would be a strange account of morality, and one which we would likely reject on philosophical grounds, rather than empirical.

Ultimately, my objection to attempts at moral debunking comes back to what I have argued previously: For the sorts of empirical data being used in these arguments to successfully threaten existing moral theories, we must understand morality as attempting to describe what is actually going on in the world, rather than arguing for how things ought to be. Otherwise, these accounts do little more than respond to the claim that, “x is the way people ought to behave.” with, “Ahh, but people do not behave that way, so you are mistaken!” It may be true that people do not behave in accord with the demands of morality, but this does not change what those demands are. Moral philosophers are not trying to describe the world, and as a result of this, their theories do not have practical consequences which can be empirically tested and refuted. Thus, as with the previous two assumptions, we must reject the assumption that prescriptive moral theories have empirically falsifiable consequences.

## 2.5 Conclusion

All three of the major assumptions which ground empirical attempts at moral inquiry cannot be adequately supported. For the empirical approach to be viable, we must believe either that philosophers do not have any particular moral expertise, that moral philosophy is a descriptive project, or that there are empirical consequences of moral theories that can be tested and potentially debunked. In the case of expertise, while philosophers may indeed not be expert *practitioners* of morality, it is much more difficult to argue that they

are not experts with regard to moral judgment or analysis. So long as this is true, we must grant that philosophers do have some special expertise which allows us to privilege their intuitions above those of non-philosophers, contrary to the assumptions supporting empirical approaches. Similarly, while the empirically-minded philosophers put forth moral theories that are ultimately descriptive, this means they are engaging in a different project than most moral philosophers, who are concerned with prescribing norms. Finally, it does not undercut prescriptive moral theories to show that people do not behave in ways they describe, or that common sense intuitions do not match up with philosophical ones. Without being able to rely on any of these three methodological assumptions, it is not clear what ground is left for the empirical ethicist, as these were necessary to make the project viable in the first place. And so, we must conclude that such approaches have relatively little to contribute to moral philosophy.

## Chapter 3

### 3 Facts and Values

How we understand the relationship between fact and value claims fundamentally shapes how we approach moral inquiry. If we understand the two to be entirely separate, then descriptive claims simply cannot be informative in moral discussions. If, however, there is no such clear divide, then descriptive statements can potentially be much more informative to moral philosophy, and the empirical approach can be on a more secure footing. This gives proponents of the empirical approach a potential line of response to the arguments from the previous chapter. They can argue that the understanding of moral philosophy I have put forward rests on a mistaken conception of the separation between facts and values. To establish that there is no is/ought gap, philosophers have offered various arguments meant to “bridge” the gap by drawing normative conclusions from purely descriptive premises. This chapter focuses on the fact/value distinction and the possibility of success for these “bridging arguments.”

This chapter has four major sections. The first section of this chapter focuses on work done by philosophers of science interested in the relationship between facts and values, coming from both feminist/standpoint approaches to philosophy of science and more traditional approaches. I argue that the concerns of philosophers of science differ from those of moral philosophers, and the account of the relationship between facts and values developed by philosophers of science is compatible with maintain an is/ought gap in moral reasoning. In section 2, I look Hume’s statement of the is/ought gap (1738 §3.1.1), and consider the meaning and scope of that argument. I argue that we are not concerned with just any argument containing apparently descriptive premises and an apparently normative conclusion. Instead, we are concerned with arguments that move from a purely descriptive set of premises to a morally substantive conclusion. In the third and fourth sections, I consider various challenges that been raised by philosophers who have put forward arguments that purport to cross the is/ought gap, focusing first on “single-premise” arguments, then on “multi-premise” ones. I argue that although some of these “bridging arguments” use only descriptive premises to validly infer a normative

conclusion, none successfully cross from a morally vacuous set of premises to a morally substantive conclusion. Since an argument would need to accomplish the latter to successfully bridge the is/ought gap, this gives us reason to continue to believe in a gap between the descriptive and the normative. I conclude this chapter by arguing that we have no reason to think that any argument can successfully move from a set of non-moral premises to a conclusion that is in any way morally substantive. I further conclude that this argumentative gap must be addressed in any argument that philosophers make that uses empirical data to generate normative conclusions.

### 3.1 Philosophy of Science and the Fact/Value Distinction

Already, one potential objection to this line of inquiry is likely to occur to defenders of a scientific account of ethics, or philosophers of science more broadly. To many philosophers of science, it will likely seem odd that I am discussing the fact/value distinction at all, given the number of arguments against its existence. Over the past 30 years, philosophers of science – led by those working in feminist philosophy of science – have strongly criticized the ideal of a value-free science championed by the more traditional approach to philosophy of science (and especially the logical positivists).<sup>59</sup> Many articles and books have been written arguing that every aspect of science – choosing research programs, theory selection, how hypotheses are articulated and tested, observations, and even what we take to be a fact – is thoroughly infused with values, both epistemic and moral. One of the major conclusions of these arguments (and the one most relevant here) is that there is no such thing as a value-free fact, and thus no sharp fact/value distinction.<sup>60</sup> But if this is the case, then the entire idea of a fact/value distinction seems to be in jeopardy. If all facts are value-laden, then how can we possibly draw such a distinction? If this is correct – and we can assume that it is for this discussion

---

<sup>59</sup> This is not to suggest that the argument against the ideal of value-free science began here. The idea that all scientific inquiry is value-laden goes as far back in the history of philosophy of science as one cares to look, and is central to the philosophy of the pragmatists, among others. However, the relatively recent argument from the feminist tradition are the ones which have had the most influence and been the most critical of the idea of a fact/value distinction.

<sup>60</sup> For an incomplete but informative list of authors who have both directly and indirectly addressed this idea, see: Harding 1986; 1991; 1993; 1995; Haraway 1988; 1989; Lognino 1990; 1993; Okruhlik 1994.

– then it seems to be problematic for the use of a fact/value distinction in ethics, and to any argument I may make that ethicists still need to be wary of the is/ought gap. How can we handle this apparent tension?

I believe that what ethicists are concerned with when they talk about a fact/value distinction and the is/ought gap is a related but different matter from what philosophers of science have concerned themselves with when they discuss the distinction. What ethicists are (or should be) concerned with is an attempt to derive normativity from a set of statements about non-normative facts. We are pushing against arguments like, “A plurality of Canadians voted for the Conservative Party in 2011, therefore, the policies of Stephen Harper are all morally above reproach.” The concern on the philosophy of science side, however, is with attempts to dismiss claims as being ‘merely’ value claims and thus uninformative to the language of science, which fails to recognize the influence values have on our judgments of facts. These two concerns seem to be compatible with each other. To see this more clearly, it will help to look at an example from Putnam.

Putnam has argued against the fact/value dichotomy, but as he points out, it is a mistake to conflate the rejection of a dichotomy or dualism with the rejection of a distinction more broadly (2002, 9). A dichotomy, as he points out, results in the belief that both sides of the dichotomy are discrete natural kinds that can be distinguished by unique essential properties that they possess (13). This idea informs his rejection of Hume’s statement of the is/ought gap on the grounds that it “assume[s] a metaphysical dichotomy between ‘matters of fact’ and ‘relations of ideas’” (14). What Putnam concludes, however, is that there is still a *distinction* to be drawn between ethical and non-ethical judgments in some limited way, but that “nothing metaphysical follows” from this distinction (19).<sup>61</sup> But if this is the position Putnam takes as his target (and the target of the philosophy of science critique more generally), then I can agree with it entirely.

---

<sup>61</sup> Putnam’s discussion on this topic continues, but what he is concerned with is the ways in which misunderstanding the nature of the difference between facts and values leads us to misunderstand the ways in which facts are value-laden. A major overall aim of the project is to show that the study of economics is fundamentally mistaken in thinking that “value questions” are outside its scope and irrelevant to its results or policy recommendations (Putnam, 2). That is, to show that “value-free” science falls flat and ignores the fundamental influence that values have on every aspect of inquiry.

Putnam argues that Hume's point is ultimately metaphysical, while I argue in the following section that it is logical, but, as I discuss below, my reconstruction is explicitly *not* meant to capture an accurate picture of the views actually held by Hume. I aim only to give an account of that argument that can still be of use to us, whether it is Hume's own view or not. The rest of what Putnam says – that there is a distinction between the descriptive and normative facts, but it is not a metaphysical one and does not represent a sharp line between the two – is entirely compatible with the position I defend.

The rejection of the fact/value dichotomy in philosophy of science, then, seems to be based on principles which the account of the is/ought gap with which I am concerned can perfectly well accept. We both agree that fact and value terms do not represent distinct natural kinds, or possess some essential property that marks them as belonging to one side or the other. So, if there is any distinction to be drawn between the two, it must stem from what we are primarily attempting to express, or a logical gap between the two. Philosophers of science have tended to focus on the influence values have on what we regard as facts, not on whether or not a sound argument can be assembled with purely descriptive premises and a normative conclusion, but this is unsurprising. It is also the case that ethicists have not often commented on the ways in which values inform all aspects of scientific inquiry. We are looking at the same problem from opposite directions, but the two pictures that arise appear to be complimentary.

## 3.2 The Is/Ought Gap

In his original statement of the is/ought gap, Hume observes a tendency of arguments of other philosophers of his time to proceed by making a series of claims about what *is* the case (i.e., descriptive claims), then conclude with a claim about what *ought* to be the case (i.e., a normative claim). He concludes that arguments of this type are necessarily deficient, as they fail to explain how “this new relation can be a deduction from others, which are entirely different from it” (§3.1.1). In this section, I will explore Hume's argument, with a focus on what force – if any – this argument has or should have in contemporary moral philosophy.

In looking at Hume's arguments, there are two ways we can potentially approach this analysis. First, we could ask what the historical figure David Hume was arguing for in this passage, or similarly, how this argument could be reconstructed to be maximally logically consistent with Hume's views more broadly. To answer this question, we would need to look into Hume's work and the *Treatise* more broadly to see how his statement on the is/ought gap fits into this broader context. This requires going into some detail on Hume's views on metaphysics, epistemology, ethics in general, the historical context of Hume's writing, and various historical interpretations that have been offered. While this could give us a fairly solid account of what Hume was attempting to accomplish with this argument, a historical account of Hume's philosophy is outside the scope of this project. The aim of this section is not to answer historical questions about Hume's views, but rather to explore the value that this argument has to contemporary moral philosophy. That Hume believed that arguments that moved from descriptive to normative claims were inadequate tells us about Hume, and may give us reason to believe there is something to the claim, it does not show that the problem he raises is one we must still handle today. Similarly, showing that Hume's statement of the is/ought gap is logically consistent with and follows from his other views does not tell us anything interesting for our present purposes. Thus, I will not be arguing about the proper historical interpretation of Hume.

The second way we can approach this – and the one which will be informative here – is to frame the question this way: is there a reasonable and consistent way we can interpret this argument that raises points which are still of interest to ethicists? Answering this question requires looking carefully at the argument itself and the various ways it can be interpreted, and which of these arguments can or cannot be successfully refuted. Importantly, it does *not* require that whatever interpretation is finally offered be consistent with other elements of Hume's philosophy or that it accurately reflect Hume's intention. As the purpose of this section is to evaluate the status of the is/ought gap, this is the more appropriate way to look into this question for our purposes. We are interested in how this argument works, not the views of Hume, and thus looking specifically at this argument without being tied down by the rest of his work best serves this goal.



The key claim of Hume's argument is that 'is' and 'ought' claims express "entirely different" relations. It is the fact that these relations are different that means moving between descriptive and normative claims does not work. In that case, what we now need to consider is what sort of relations we are concerned with here, and whether or not the two types of claims do in fact represent different relations. It is important to bear in mind that Hume is talking about problematic argument structures and inferences here. While this may seem trivial, it is nonetheless worth noting. What he is commenting on is not any problematic premise or set of premises which he sees commonly arising, but with the ways in which people move from acceptable premises to a conclusion which seems to be a non-sequitur, or at least questionably related to the premises. What this tells us is that Hume is making a *logical* point here. The argument is not that the philosophers he is discussing are getting the metaphysics of value claims wrong, or anything else that commits us to any sorts of beliefs about ontology. Rather, the claim is that "x is y" describes a different logical relation than "x ought to y", and further, that these relations are such that we cannot cross from the former to the latter without some additional premise.

Intuitively, it is easy to see the appeal of this idea. One of the earliest things any undergraduate philosophy student learns about logic and arguments is that an argument which introduces a new term in its conclusion is (generally) going to be invalid. Given this, it seems natural to think that the introduction of an "ought" in the conclusion of an argument which has relied entirely on "is" premises up to that point marks the argument as invalid, if it is in fact the case that the two terms express entirely different logical relations. Still, we must get clear on exactly what sorts of arguments we are and are not concerned with here. If we are saying that there exists absolutely no logically valid or sound argument which contains exclusively descriptive premises and a normative conclusion, then even a basic understanding of the rules of logic will show that it is false. For instance, a tautology follows from any possible sentence or set of sentences. Thus, if the conclusion of our argument is a normative one that is also tautological, trivially we can derive it from a set of purely descriptive premises. Similarly, since any sentence whatsoever is derivable from a contradiction, we can quite easily generate any number of arguments that move from descriptive premises to a normative conclusion by having as

its premises any descriptive statement and its negation, and having any normative sentence at all as its conclusion.

If we are to be at all charitable to Hume's argument, we should resist interpreting it as meant to capture cases that are uninformative or trivial. Instead, what we must be concerned with are arguments in which a) the descriptive premises are related somehow to the normative conclusion, and b) the normative conclusion is in some way a substantive one. I suggest the following as a test of whether or not a sentence is morally substantive: Sentence  $x$  is morally substantive iff  $x$  is a sentence of language  $y$ , and a competent but morally ignorant speaker of  $y$  could learn a new moral fact from  $x$ . In other words, for a sentence to be morally substantive, it must be possible to learn something moral from it. So, a sentence like, "It is wrong to commit murder" would be morally substantive, while, "It is not the case that it is wrong to steal and it is not wrong to steal" would not be. In the former case, a person who is entirely ignorant of the demands of morality could learn something new if the statement is true. In the latter, what is expressed is equivalent to  $\sim(P \wedge \sim P)$ , and thus is not potentially informative to anyone who already knows what all of those words mean. If the Humean picture sketched above is correct, there can be no argument that moves from purely descriptive (or morally insubstantive) premises to a morally substantive conclusion.

### 3.3 Single-Premise Arguments

There have been many attempts to show, contra Hume, that normative conclusions can be derived from purely descriptive premises. Various philosophers have presented a number of arguments which they take to have purely factual premises and normative conclusions, employing a number of different logical or linguistic approaches to do so. For our present purposes, I focus on the accounts of Prior (1960), Searle (1964), Geach (1977), and Nelson (2007). I argue that, while these arguments require us to get a clearer account of what marks something as a normative or non-normative claim, none have successfully managed to present a sound argument consisting of purely descriptive premises and a morally substantive conclusion. First, let us consider what I am calling, "single-premise"

arguments – which is to say, arguments consisting of a single, unconjoined premise such that we could reasonably represent them in first-order logic as a single atomic sentence.<sup>62</sup>

There are three major example arguments put forth by Prior, but for now I will look only at the first two, and come back to the third later, as it has much in common with Geach's approach, and thus is mostly usefully considered alongside it. His first argument (**A1**) has as its premise, "Tea-drinking is common in England", and concludes, "therefore either tea-drinking is common in England, or all New Zealanders ought to be shot" (201). The argument is undeniably valid (it is a straightforward example of addition), and the sole premise is plainly descriptive (and more importantly, plainly non-normative). It is not quite so clear that the conclusion is a normative one. It contains the word "ought", to be sure, but the sentence as a whole is a disjunction containing a non-normative disjunct. Prior addresses this by presenting this in the context of a different, more traditionally normative argument (**A2**): "[P1] Anyone who does what is not common in England ought to be shot; [P2] All New Zealanders drink tea; therefore [C] either tea-drinking is common in England or all New Zealanders ought to be shot" (201-2, premise and conclusion numbering mine). With this formulation, it is easier to see how the conclusion could be taken as a normative claim. This follows a format we recognize as being one which leads to normative conclusions, and so, Prior concludes, since the sentence "either tea-drinking is common in England or all New Zealanders ought to be shot" turns out to be a normative one when we see it in **A2**, it is also normative in the earlier case, and thus **A1** successfully moves from a single descriptive premise to a normative conclusion.<sup>63</sup>

In order for this argument to work (that is, for it to successfully cross the is/ought gap), it must be the case that the conclusion of **A1** is actually normative. That is, Prior must assume, as Scott Hill points out, "that a sentence's classification as ethical or non-

---

<sup>62</sup> For example, "It is raining" would count as a single premise, while "it is raining and it is Wednesday" would not.

<sup>63</sup> Prior also presents several shorter arguments, but they are not different enough from this one to be worth discussing in any detail.

ethical is invariant and context independent” (2008, 547). If it is, then the fact that the sentence seems clearly normative in its second use is relevant. If the classification can change based on context, however, then pointing out that it is a normative claim in a different context is entirely useless. Hill argues that disjunctive claims of the type Prior uses are context-dependent by using the taxonomies of Karmo (1988) and Maitzen (1998) for distinguishing ethical from non-ethical statements<sup>64</sup> to show that the conclusion is non-ethical in the **A1**, despite being ethical in **A2**,<sup>65</sup> thereby establishing that Prior has not moved from purely descriptive premises to a normative conclusion.

While I do not disagree with Hill that both Karmo and Maitzen’s taxonomies demonstrate a flaw in Prior’s argument, the use of the taxonomies does not appear to me to be the clearest way to demonstrate this. Instead, it may be more helpful to consider the ways in which we actually use disjunctive claims. There are two major ways in which disjunctive statements tend to be used in conversation.<sup>66</sup> The first is to express uncertainty between two possibilities, with the disjuncts being related to each other in some (e.g., “The package will arrive either today or tomorrow.”). The second is generally rhetorical, with one disjunct being a proposition in which we have and are expressing extreme confidence, and the other being something outlandish or irrelevant, often with the second disjunct being a non-sequitur (e.g., “Either J.S. Mill wrote *On Liberty*, or I’m the King of France.”). The former is expressing an equivalent to a conditional of the form, “If not P, then Q”, while the second is just a strong assertion of P.

When we look at Prior’s two arguments bearing these two understandings of disjunctions in mind and considering the pragmatics of the conclusions, it becomes clear

---

<sup>64</sup> The taxonomies break down as follows: For Karmo a sentence S is “ethical in a possible world w just in case S is true in w with respect to one ethical standard, and false in w with respect to another ethical standard” where an ‘ethical standard’ is understood as a standard which assigns a truth value to all uncontroversially ethical sentences (254). For Maitzen, “ $\phi$  is an ethical sentence only if it is impossible for  $\phi$ , standardly construed, to be true unless a given moral property (such as moral rightness or wrongness) is possessed by at least one object” (361).

<sup>65</sup> Karmo also shows how his taxonomy arrives at this judgment (256); Maitzen does not. As Maitzen is concerned with demonstrating that the gap *can* be crossed, it is unsurprising that he does not address it.

<sup>66</sup> I am concerned, of course, with the pragmatics of disjunctive claims here, not the semantics.

that the conclusion is expressing something different in each case, and only one of them is normative. In **A2**, the “speaker” of the argument is unsure as to whether or not the first disjunct (“tea-drinking is common in England”) is true, and whether or not tea-drinking is common in England determines whether or not all New Zealanders ought to be shot, according to this argument. So, in this case we can, without sacrificing anything in the pragmatics of the sentence, take the sentence as saying, “If tea-drinking is not common in England, then all New Zealanders ought to be shot.” This is clearly a moral claim, as we can see if we consider structurally and pragmatically similar sentences like, “If anyone is not a vegetarian, then her diet is morally wrong”, or “If Dennis commits murder, he is evil.” All three express a commitment to a particular norm or set of norms, and some sort of connected moral judgment.

The conclusion of **A1**, however, seems to fall into the pragmatics of the latter kind of disjunction. The conclusion is true because the first disjunct is simply a restatement of the premise of the argument. If we’re concerned with the pragmatics of the conclusion, then, we cannot take this utterance of the conclusion as being the same as the **A2** utterance. Instead, what is being expressed is something closer to, “I strongly believe that tea-drinking is common in England, and believe it strongly enough that I’m willing to say that something ludicrous would be the case were it not true.” But this is simply an emphatic statement of a purely descriptive claim. By treating the two instances of the conclusion as serving the same function, Prior fails to recognize the importance of context in determining a sentence’s meaning, and thus fails to recognize that its status as descriptive or normative changes along with that context. So, since the use of the conclusion in **A1** is not normative (despite having the semantic form of a potentially normative claim), this argument does not move from descriptive premises to a normative conclusion and fails to cross the is/ought gap.

Prior’s second argument is similar, but there are a few key differences that make it worth briefly bringing up. Here, the single premise, “Undertakers are Church officers” leads to the conclusion, “*whatever* all Church officers ought to do, undertakers ought to do” (203, original emphasis). The key difference between this and the first case is that, as Prior notes, the conclusion is not what he calls “contingently vacuous” (204). In the first

example, the “ought to” in the conclusion can be replaced by anything else that fulfils that grammatical role with no impact to the argument (since the first disjunct is the one doing all the work). In this case, however, the first “ought to” cannot be replaced with any similarly grammatical set of words. If we subbed in, for example, “think they ought to” in its place, the conclusion would no longer follow. In other words, unlike the first case, it is not just an odd feature of the conclusion that it happens to be normative. Rather, the normative aspect of it is central to the entire argument. This may help with the issues discussed above, as it means that the normative part of the conclusion serves as more than merely a pragmatic emphasis tacked on to a descriptive claim. This conclusion results in a conclusion that could give potential normative advice (presumably to an undertaker unsure of her duties).

As in the previous case, though, it is not clear that this example actually does manage to cross the is/ought gap. Karmo argues that, by his taxonomy, one of two things will be the case: either the conclusion is non-ethical, or the argument is unsound (256). To evaluate the argument by his standard, there are two possible worlds we need to consider: one in which the premise is true (i.e., the world in which undertakers are in fact Church officers), and one in which it is false. In the latter, the sentence is ethical, as we can craft two standards: one in which the claims about moral duties of undertakers match up with the claims about moral duties of Church officers, and one in which Church officers have at least one duty that undertakers to not. As the premise is false in the latter world, however, the argument is not sound. Conversely, in the possible world in which undertakers are Church officers, there can be no consistent ethical standard by which the claim is false, as any moral claim that applies to the class of all church officers must also apply to everything which is a member of that class, including undertakers. In that case, the conclusion is non-ethical, and thus does not cross the is/ought gap.

While I agree that this argument works, I share Hill’s concern that Karmo’s taxonomy gets the wrong answer on several important classes of sentences, including necessary claims (Hill 2008, 563ff.). For Karmo, any necessary claim will always come out as non-ethical, as there can be no consistent assignment of truth values which falsifies it. It seems, however, that when we utter a moral tautology such as, “All morally wrong

acts ought not to be done,” or, “All of my moral duties ought to be performed”, we are making ethical claims.<sup>67</sup> If this is the case, and moral tautologies *are* properly considered to be normative claims, then Prior’s example is in better shape. Given his premise, the conclusion of the argument becomes, “All members of a class are morally obligated to do whatever is morally obligatory for the class as a whole.” This statement is tautological, or at least uninformative (it only really says that whatever is obligatory for x is obligatory for x), but it will also turn out to be normative, in which case, the argument is a sound one which does move from descriptive premises to a normative conclusion.

Despite this, however, we should still not consider an argument of this sort to be a successful bridging of the is/ought gap, or at least not an interesting one. If the normative claim at which we arrive in the conclusion is tautological, or not morally substantive, then as I said above, it is no surprise at all that we can derive it from descriptive premises. We can also derive it from normative premises, false premises, or no premises whatsoever. This is not getting us to the sort of substantive normative conclusions we would want to see in order to count a bridging argument as managing to show that there is no is/ought gap. Prior’s conclusion is – in any possible world in which undertakers are Church officers – vacuous, and thus we need not concern ourselves with it.

A similar case comes from Mark Nelson in the form of the argument:

1. Stalin authorized the Katyn Massacre.

Therefore,

2. It is not the case that both *If Stalin authorized the Katyn Massacre, then Stalin was evil* and *It is not the case that Stalin was evil*. (Nelson 214)

---

<sup>67</sup> Admittedly, my claim here relies entirely on my own intuition (which I share with Hill) that such claims are in fact normative ones, but any attempt to craft a taxonomy of ethical statements will require at least some use of intuition regarding how certain sentences are to be classified.

The single premise of the argument certainly appears to be descriptive, and Nelson argues that the conclusion is normative by the standards of Karmo's taxonomy (Nelson 214).<sup>68</sup> But as in Prior's example, the conclusion is normatively hollow, and thus not a threat to the is/ought gap. To see why this is the case, consider the logical form of the conclusion, which is  $\sim[(P \rightarrow Q) \wedge \sim Q]$ , or  $\sim(P \rightarrow Q) \vee \sim(\sim Q)$ . So, for this statement to be true, the conditional " $P \rightarrow Q$ " must be false, or  $Q$  must be true. For  $\sim(P \rightarrow Q)$  to be true, it must be the case that  $P$  is true, and  $Q$  is false, so, the conclusion becomes  $(P \wedge \sim Q) \vee Q$ . Now, given the premise, we are concerned only with possible worlds in which  $P$  is true (as that is the sole premise). In that case, however, the conclusion will be true regardless of the truth value of  $Q$ . If  $Q$  is true, the second disjunct is satisfied; if it is false, the first is satisfied. In that case, in the only possible worlds in which we pay any attention at all to this argument, the conclusion just says, " $Q \vee \sim Q$ " which is a clear tautology.

We are not being told anything non-trivial by this argument beyond a mere assertion of the premise. To see this more clearly, note that we can sub in any true premise – including ones that are completely irrelevant to the normative component of the conclusion – and this argument will lose none of its force. The argument, "Today is Saturday, therefore it is not the case that both *If today is Saturday, then Stalin was evil* and *It is not the case that Stalin was evil*" is also valid, but surely we do not think that the day of the week is in any way normatively informative. Similarly, any sentence can be subbed in for  $Q$ , including complete gibberish, and the argument holds. As before, the argument, "Stalin authorized the Katyn Massacre, therefore it is not the case that both *If Stalin authorized the Katyn Massacre, then colourless green ideas sleep furiously* and *It is not the case that colourless green ideas sleep furiously*" is also valid, but it tells us nothing of interest about the sentence in question. Nelson's argument may initially appear as if the addition of the normative clause in the conclusion is related to the premise, since the relevant sentences both concern a single subject (i.e., Stalin), but looking at the logical structure of the argument, it becomes clear that there is no relation between the

---

<sup>68</sup> Hill (2009) argues that Nelson is mistaken, and the conclusion is in fact non-normative, but I am willing to grant Nelson that the conclusion is a normative one. As we shall see, it makes no difference.



descriptive premise and the normative clause of the conclusion. Like any single-premise argument, the only conclusion that can be generated is some combination of that premise and a tautology. But, of course, tautologies are uninformative, so their addition to any argument does not actually add anything to the already-established premises. All Nelson has demonstrated is that we can show that normatively tautological claims follow from any true premises, but this does not show that any interesting normative claims can be derived from purely descriptive premises, which is what we would need to see to consider an argument to be a counterexample to the is/ought gap.

### 3.4 Multi-Premise Arguments

A different approach comes from Prior's third major argument, and ideas put forward by Geach. Over the course of three papers,<sup>69</sup> Geach lays out a number of arguments that he believes manage to derive normative conclusions from descriptive premises, one of which is strongly similar to an example given by Prior. Prior's arguments is as follows:

No one ought to do what invariably accompanies the doing of something wrong;

X.Y. invariably acts as he says he will act;

Therefore X.Y. ought never to say that he will do anything that he ought not to do.  
(206)

Geach's similar argument concerns a promise made to a dying uncle:

- (1) If Evan were to promise Uncle Dewi to adopt some practice, then he would adopt it.
- (2) If Evan were to utter sentence W in the case described, he would be promising Uncle Dewi to adopt the practice of doing something he ought not to do at least twice every day.
- (3) Nobody ought to adopt the practice of doing something he ought not to at least twice every day.

---

<sup>69</sup> Geach 1976; 1977; 1982

[. . .]

- (4) If Evan were to utter sentence W in the case described, he would adopt the practice of doing something he ought not to do at least twice every day.

[. . .]

- (5) Evan ought not to utter sentence W in the case described. (Geach 1977 475-6)

Both of these arguments concern the same issue: that a description of a given speech act seems to commit us to a normative conclusion. However, as both of these examples use claims that are normative (if vacuous) by the standards I have suggested above, it will be more useful to consider an example in the style of these two arguments put forward by Karmo:

'Everything that Alfie says is true' looks like the sort of sentence one should call non-ethical; 'Alfie says that it ought to be the case that everyone is sincere' is non-ethical; but from these there follows the sentence 'It ought to be the case that everyone is sincere', which is ethical. (Karmo, 253)

This example has the advantage that it keeps the part of Prior and Geach's examples that is of interest (it is not a single-premise argument, as the earlier cases were, and it moves from a description of a speech act to a normative conclusion), but does not contain any *prima facie* normative premises, vacuous or otherwise. So, let us consider this argument, and see if it does in fact move from descriptive premises to a normative conclusion.

As before, Karmo analyzes the argument in terms of his taxonomy, and argues that in any possible world in which the argument is sound, the first premise will turn out to be an ethical one (255-6). Although, once again, I agree that his taxonomy can establish this, it seems we can accomplish the same goal by simply looking at the logic of the premises, without needing to defend any particular taxonomy of moral and non-moral statements. The first premise of the argument is, "Everything that Alfie says is true", which Karmo describes as "look[ing] like" a descriptive claim (253), and I said was "prima facie" not normative. The sentence certainly initially appears to be an "is"

sentence,<sup>70</sup> and thus seems to be purely descriptive, but this begins to collapse when we bear in mind the logical form of universal claims. A universal claim can be considered a series of conjunctions with each conjunct being one of the members of the class specified by the universal and the claim that the member possesses the property in question. In this case, the class is ‘sentences uttered by Alfie’ and the property is ‘truth’. The claim, “everything that Alfie says is true”, then is not a single claim, but rather a conjunction of each individual statement uttered by Alfie, with each one followed by, “is true.” We know from the second premise that one of the sentences Alfie has uttered is, “It ought to be the case that everyone is sincere.” In that case, the first premise – properly unpacked – reads, “Claim one Alfie made is true, and claim two Alfie made is true, and ‘It ought to be the case that everyone is sincere’ is true, and claim three...” and so on, until we run out of statements uttered by Alfie. But in this case, our first premise must be at least partly normative, since it contains within it an endorsement of a moral claim.<sup>71</sup> In that case, while the conclusion of the argument may indeed be a moral one, it is simply not the case that the premises are purely descriptive.

This brings us to Searle’s argument, which I take to be the most interesting approach. Searle presents the following argument:

- (1) Jones uttered the words "I hereby promise to pay you, Smith, five dollars."
- (2) Jones promised to pay Smith five dollars.
- (3) Jones placed himself under (undertook) an obligation to pay Smith five dollars.
- (4) Jones is under an obligation to pay Smith five dollars.
- (5) Jones ought to pay Smith five dollars. (Searle 44)

He further supplements these with the premises:

---

<sup>70</sup> That is, after all, the verb in the sentence.

<sup>71</sup> This operates on the assumption that “‘x ought to y’ is true” does not express anything importantly different from “x ought to y”, but in the context of anything other than a debate on whether or not normative claims have truth values, this seems uncontroversial enough.

(1a) Under certain conditions *C* anyone who utters the words (sentence) "I hereby promise to pay you, Smith, five dollars" promises to pay Smith five dollars. [. . .]

(1b) Conditions *C* obtain. [. . .]

(2a) All promises are acts of placing oneself under (undertaking) an obligation to do the thing promised. [. . .]

(3a) Other things are equal. [. . .]

(3b) All those who place themselves under an obligation are, other things being equal, under an obligation. [. . .]

(4a) Other things are equal. (44-6)

One method of arguing against Searle has been to argue that inference from (4) and (4a) to (5) is either invalid, or rests on a normative (or “evaluative”, to use Searle’s terminology)<sup>72</sup> sense of “other things are equal” being employed.<sup>73</sup> This avenue of discussion has been thoroughly explored, and I do not wish to address it further here. Instead, I want to consider a different approach to this argument that has also been brought up by Searle himself.

In discussing potential objections to his position, Searle raises two that are of interest. His first concern is the objection that one of his premises tacitly smuggles in normative talk (49-50). He considers two possible places that he could be accused of doing this. The first concern is that “there must be a concealed evaluative premise in the description of the conditions in (1b)”<sup>74</sup> (49), which he rejects, claiming that it is either question-begging, or his interlocutor must show exactly how conditions *C* unavoidably

---

<sup>72</sup> For the duration of the discussion of Searle, I will stick with his use of “evaluative.” There does not appear to be a difference between Searle’s use of this term and “normative” as I have been using it, so far as I can tell.

<sup>73</sup> Cf. Thomson & Thomson (1964), Hill (2008 547-552).

<sup>74</sup> i.e., the conditions that must obtain for the utterance to count as a promise.

contain non-descriptive claims (49-50). While it is not clear that the objection does in fact beg the question any more than Searle's own assertion that *C* does *not* necessarily need to include evaluative claims,<sup>75</sup> Searle is nevertheless correct that *C* does not need to include anything but descriptive claims in order to be an account of the conditions under which a given speech act successfully convey that the speaker has made a promise.

Consider the set of circumstances that Scanlon claims constitute having made a promise according to his Principle F:

If (1) A voluntarily and intentionally leads B to expect that A will do X (unless B consents to A's not doing so); (2) A knows that B wants to be assured of this; (3) A acts with the aim of providing this assurance, and has good reason to believe that he or she has done so; (4) B knows that A has the beliefs and intentions just described; (5) A intends for B to know this, and knows that B does know it; and (6) B knows that A has this knowledge and intent; then, in the absence of special justification, A must do X unless B consents to X's not being done. (Scanlon 1998 304)

This principle contains normative language in its consequent, but we are not concerned with Scanlon's account of the binding nature of promises. What is relevant here is conditions (1) through (6). Searle can easily say that exactly these circumstances are the ones that make up *C*. All of them are fairly clearly descriptive, and match up well enough with our common sense intuitions about what constitutes a promise.<sup>76</sup> In this case, Searle can give a clear account of *C* that contains no evaluative terms, and that objection is not a concern.

Searle also looks at a variant of this objection which claims that the argument only really shows that promising is an evaluative concept. This, he claims, will not work, since if that is correct, the seemingly acceptable inference from (1) – a mere description of a speech act – to (2) – the speaker has made a promise – has already moved from

---

<sup>75</sup> Searle gives a list of some potential conditions (44-5), but he does not argue that there need be no further ones, or that a full description of *C* will not include evaluative premises.

<sup>76</sup> This is not to say that Scanlon's account is unproblematic, but an evaluation of the merits of Scanlon's account of promising is well beyond the scope of this project. All I am claiming here is that Scanlon can avail himself of Principle F to show that *C* need not contain anything normative to be a plausible account of the circumstances under which a given speech act functions as a promise.

descriptive to evaluative, and thus he has accomplished what he set out to do (50). While Searle separates this from the second objection he considers, what is centrally at issue is very similar to what matters to the second objection. Given this, let us turn to the second point and consider them together.

The second objection Searle considers is that the core principle of the entire argument, “one ought to keep ones promises” is a moral/evaluative claim, and thus the argument does not move from descriptive to normative (50-1). He responds that, if it is moral, it is tautological, since it derives from the tautologies, “All promises are (create, are undertakings of, are acceptances of) obligations,” and, “One ought to keep (fulfill) one's obligations” and proceeds to diagnose why philosophers have failed to appreciate the tautological nature of this claim (50-1). This defense does not help Searle’s position. Plainly, one of two things is the case: either “one ought to keep ones promises” is tautological,<sup>77</sup> or it is not. If it is not, then the derivation rests on a tacit normative premise, and Searle has not derived “ought” from “is”. If it *is* tautological, that must mean that its two constitutive premises are as well (as Searle asserts). If *this* is true, then “obligation” must be analyzable as (roughly), “thing one ought to do.” In this case, premise (2a) can be reworded with no change in meaning to (2a’), “All acts of promising make it such that one ought to do the thing promised,” which can be broken down logically to, “For any x, if one promises to do x, then one ought to do x.” Both of these sentences are plainly normative. This means that Searle’s premise (2a) must – by his own understanding – be a tacitly evaluative one. In that case, the move from (2) to (3) relies on normative claims, and the argument is not an entirely descriptive one.

This brings us back to the possibility that “promise” is itself an evaluative term. Searle says that, if it is, the move from (1) to (2) is a move from descriptive to evaluative. This also is not entirely correct. If “promise” expresses something evaluative, then for the move from (1) to (2) which explains that the speech act constitutes a promise cannot contain only the descriptive element of what a “promise” is; it must also have some

---

<sup>77</sup> And we will provisionally grant Searle’s assumption that the presence of tautological evaluative claims is fine

reference to the evaluative component of the term. Again, this does not represent a derivation of evaluative conclusions from descriptive premises.

Taken together, all of this raises a new question: if terms like “promise”<sup>78</sup> or “obligation” can carry normative content, what are we to make of sentences like, “Dee promised to pay Mac five dollars,” or, “Charlie is obligated to give to charity”? These have characteristics of descriptive sentences, but seem to also make evaluative claims. Similarly, the previously discussed claim, “Everything Alfie says is true” has evaluative components assuming he utters at least one normative phrase, but if he also utters at least one descriptive phrase, it also certainly has descriptive components. So, how ought we to classify these phrases as a whole? Are they descriptive or normative? Although his argument does not ultimately work, the overall conclusion Searle draws is still quite informative here. Searle believes he has shown that his argument crosses the is/ought gap, and sets out to explain why philosophers had previously believed that doing so was impossible. Searle’s diagnosis is roughly this: there are prototype descriptive sentences (“It is raining”) and prototype evaluative sentences (“Murder is wrong”). These are clearly different types of sentences, with different standards for evaluating them,<sup>79</sup> which leads to the belief that there is a sharp line between the two. What Searle thinks this misses are what Searle calls, “institutional facts”, which exist only in light of some institution or convention (for example, “Bishops in chess can only move diagonally”) (53-5).

Searle claims that such facts are a different type of descriptive claim from which one can derive evaluative consequences (57-8). So, for his example, because we have an institution of promising, and part of that institution is that promises are to be kept, we can move directly from “x promised to do y”, which Searle takes to be descriptive, to “x ought to y”, which he does not. He concludes that we can – at least within the scope of

---

<sup>78</sup> Not that I wish to necessarily commit to terms like “promise” being normative terms, but I am willing to entertain the possibility.

<sup>79</sup> Searle claims that the difference is that evaluative claims are not capable of being true or false (53). I do not want to comment on the truth value of moral claims here. The relevant part of Searle’s claim is that there is a noticeable difference between the two.

institutional facts – derive “ought” from “is” (57). While not entirely off-base, it does not appear to me to be the best answer to the question of how we ought to understand sentences that seem to be both descriptive and normative. Instead, a more useful account will arise from bearing in mind what we saw from the single-premise arguments above. The mistake Searle has made is the same one that is made by both Prior and Nelson in our earlier discussion. Namely, assuming that a given sentence will be classified as exclusively either descriptive or normative, and that this classification does not vary by context. What we saw from the earlier examples, however, is that which way we interpret these sentences depends on the context of their use, and the intention behind the utterance. The sentence, “You promised to pay Frank five dollars” would be descriptive when being uttered in order to remind a friend of an event they had forgotten. If it is being uttered to scold that same friend when they – despite being fully aware of the promise – have decided not to fulfill it, it is a normative claim.

### 3.5 Conclusion

All of this brings us back to the central question of this chapter: What is the status of the is/ought gap? Fittingly, the answer to this question also depends on exactly what is being asked. If we are concerned with a supposed sharp line between statements such that we can categorize any given sentence as being purely descriptive or purely normative, then no such line exists. As the various arguments which have attempted and failed to derive normative conclusions from purely descriptive premises have shown us, some sentences which appear *prima facie* descriptive will in fact be employed to express a norm, and some *prima facie* normative statements will be used with a primarily descriptive function. This is something we need to be aware of as we proceed, and we must be careful not to ensure that when we encounter a *prima facie* descriptive or normative sentence, it is not being used differently than we expected.<sup>80</sup> On this understanding there seems to be no reason to think there is a sharp is/ought gap.

---

<sup>80</sup> Or relying on tacit normative premises.



If, however, we are concerned with whether or not anything morally substantive can be derived from purely descriptive premises, then the is/ought gap is still a concern. The key here is that what we are concerned with are *purely* descriptive premises. There are two ways to understand what the word “purely” means here. First, we could assume these are premises which are not partly normative, and do not admit to normative interpretation, regardless of use.<sup>81</sup> This is a slightly stronger interpretation than is both necessary and useful, however. The more useful way to understand this is as referring to an argument in which all of the premises – regardless of how they could be interpreted in other contexts – are used in the argument in a way that makes their use entirely descriptive. As we have seen from the discussion in this chapter, no one has yet managed to successfully put forward an argument that accomplishes this. Of course, the inability of anyone to present an argument that crosses the is/ought gap is not itself proof that it cannot possibly be done. However, since we already have a positive argument for the existence of the is/ought gap from Hume and from our understanding of logic more generally, combining this with our inability to find counterexamples gives us reason to believe that there is a logical gap between descriptive and normative claims, and thus we should be wary of any argument which purports to move from a description of the world to a morally substantive conclusion.

What we can say about the is/ought gap overall at this point, then, is this: We have seen that the is/ought gap, understood as a sharp metaphysical or ontological distinction between kinds of sentences, does not exist in any important way. We have further seen that, although we can make some sort of distinction between sentences which mainly aim to describe the world, and sentences which aim to give some sort of moral instruction, we must carefully examine exactly what a sentence is meant to communicate in order to determine if it is primarily making use of descriptive or normative features. Where a gap exists between descriptive and normative claims, it is an explanatory gap, or one in inferences we are able to justifiably draw from descriptive claims. An argument

---

<sup>81</sup> To some extent, any declarative statement can be interpreted as asserting an epistemological norm (by declaring sentence S, I am also declaring, “You have reason to/ought to believe S”), but as previously stated, we are interested in more robust/informative norms than these.

which lacks anything of normative force in its premises is not going to be able to draw any substantive or normative conclusions. Further, we have seen that this understanding is compatible with the rejection of elements of the fact/value dichotomy that has gone on in philosophy of science. Given this, we are left to conclude that the is/ought gap, at least as a property of the structure of arguments, remains a live problem in moral philosophy. Proponents of an empirical approach to ethics may not, therefore, defend their approach from methodological objections by replying that the idea of a fact/value distinction is irrelevant, and thus it is entirely possible to derive morally substantive conclusions from non-moral premises. The remaining options are either to demonstrate that the sorts of empirical claims made by these philosophers are morally substantive, or to admit that their conclusions fail to be morally substantive.

## Chapter 4

### 4 Companions in Guilt Arguments

At this point, one would be forgiven for thinking that the prospects for continuing to do moral philosophy within an empirically-informed framework are bleak. Data from cognitive science and evolutionary biology undermine our confidence in the intuitions that necessarily serve as a foundation for traditional moral inquiry. The most immediate alternative, the empirical approach, is grounded in methodological assumptions about ethics that are shaky at best, and unable to answer the sort of substantive moral questions we continue to ask and for which we want answers. The best hope for the empirical approach is a rejection of the fact/value distinction, and an insistence that we can derive substantive normative claims from descriptive facts. Unfortunately, no successful defense of this idea - nor any apparent examples of interesting arguments that successfully derive substantive normative conclusions from purely descriptive premises - appears to be forthcoming, despite decades of attempts.

Faced with the apparent failures of the dominant approaches to ethics, we are left with two options. We can abandon the normative project of ethics, declaring it impossible to perform in light of what our best science tells us about our evolutionary histories and basic facts about our psychology. This effectively amounts to claiming that science has shown that we cannot give any meaningful or authoritative answers to questions of what we ought to do, or that such a question is fundamentally flawed to begin with.<sup>82</sup> Our other option is to find a methodology for moral philosophy that is able to avoid the issues facing both the traditional and empirical approaches. The former option is untenable. Practically speaking, we are not going to stop wondering about how we ought to behave or what morality requires of us. As Christine Korsgaard has argued, normativity is a fundamental part of humanity, and “[h]uman life [. . .] is pervaded through and through

---

<sup>82</sup> This also practically amounts to endorsing the empirical approach. If there is no possibility of a successful prescriptive moral project, then the descriptive understanding advanced by philosophers like Prinz becomes not a misunderstanding of the project of moral philosophy, but rather the only viable option.

with obligation” (Korsgaard 1996, 255), both moral and non-moral. Rejecting this question amounts to rejecting a fundamental element of human experience. Similarly, responding to the question, “Is it okay to eat meat?” by pointing out that more people in North America eat meat than do not fundamentally fails to answer the question. So too does research that shows how we feel about animals and about eating different kinds of them. As such, the position that there is nothing meaningful philosophers (or anyone else) can say in response to moral questions is ultimately not a satisfactory route. This is not to say that there is no circumstance in which we should accept the notion that we are unable to know the answers to moral questions, or that those questions are fundamentally misguided. At a certain point, and after a certain amount of failure, we must admit that even the most persistent of questions may be unanswerable. Such a course should however, be adopted only after we have exhausted all other possibilities, which is certainly not currently the case for the study of ethics.

The question we are left with, then, is a refined version of the one with which we began. Our question as the outset was, how do we approach moral philosophy from within a naturalist framework? Now, the question is, what can we learn from the challenges facing both the traditional and empirical approaches to moral philosophy, and how can we carve out a space between these two positions? To some extent, the answer to this question is straightforward: the traditional approach struggles with being able to adequately respond to empirical data, while the empirical approach struggles with having any real normative force. Our aim, then, is a method for ethics which is properly responsive to empirical data, but nevertheless has normative force and is not merely a description of people’s reactive attitudes, or a statement about what most people think is morally right. What this actually means, however, and what such a moral methodology would look like, is much more difficult, and is the question I answer in both this and the following chapter. As these chapters involve similar subject matter, it is worth discussing the layouts of both of them together.

In this chapter, which has two sections, I look at the ‘companions in guilt’ arguments advanced to try to defend normative reasoning in moral philosophy by arguing that such reasoning is indispensable in epistemology. Although I conclude that the simple

analogy to epistemic norms does not necessarily “save” moral normativity, I nevertheless argue that companions in guilt-style reasoning can help us to better understand how moral philosophers can respond to empirical criticisms – especially those surrounding the use of intuitions – by looking at how epistemologists have done so. The fifth and final chapter continues and extends this analysis. In the first section of that chapter, I look at responses to empirical criticisms that have been offered by epistemologists, and how these could be applied to moral philosophy. While there is not a perfect analogy between the two fields such that arguments that are successful in epistemology can be applied directly to moral philosophy, there is enough similarity between these two heavily normative fields that we can begin to see potential lines of defense for moral normativity. Sections two through five reconsider the role of intuitions in moral philosophy and highlight possible understandings of our intuitions that can avoid the criticisms discussed in the first chapter. In the sixth section, I discuss what all of this means, and how it can all be brought together to give us at least some idea of how to engage in empirically-informed moral philosophizing without sacrificing normativity. I conclude that the prospects for the future of moral philosophy from within a naturalist framework are not nearly as bleak as they may appear, and that we have all the methodological tools we need to continue doing moral philosophy without giving up any of our commitments to the well-confirmed theories of our best science.

## 4.1 Defending Companions in Guilt Arguments

One strategy to counter the arguments of moral error theorists has been the companions in guilt argument (or CIGAr)<sup>83</sup>. Often attributed to Cuneo (2007), this argument holds that error theorists cannot coherently reject the categoricity or authority of moral norms without also rejecting epistemic norms, which they are unwilling to do.

Christopher Cowie presents the following general form of the CIGAr:

---

<sup>83</sup> I owe my use of this acronym to Clipsham (forthcoming).

1. *Parity premise*: if the arguments for the moral error theory are sufficient to establish its truth, then those arguments (or appropriate analogues of them) are also sufficient to establish the truth of the epistemic error theory.
2. *Epistemic existence premise*: the epistemic error theory is false.
3. So, the arguments for the moral error theory are not sufficient to establish that the moral error theory is true. (Cowie 2014, 408).

The CIGAr works by arguing that the problematic features of moral norms and reasoning that error theorists identify are also features of epistemic norms and reasoning. This presents the error theorist with a dilemma: either admit that epistemic reasoning is as problematic as moral reasoning and jettison epistemology, or admit that these features are not, in fact, inherently problematic, which readmits moral reasoning.

The particular forms of the CIGAr can vary somewhat, mainly in how they establish the parity premise. Generally speaking, the premise is established in a manner similar to Richard Rowland's argument for it. As he constructs the argument, the parity premise is established by two premises: 1) "According to the moral error theorist, moral reasons are metaphysically problematic because they are categorically normative" and 2) "epistemic reasons are categorically normative" (Rowland 2016, 161). And as Cowie points out, defending moral facts by appeal to the CIGAr may require showing that certain relevant features of epistemic norms are shared by moral norms (2014, 410). The epistemic existence premise – which we will return to later in the chapter – is generally defended by arguing that epistemic error theory is at best so deeply unappealing that it ought to be rejected, and at worst simply self-defeating.

It is worth taking a moment here to note that, although in this section I will largely be talking about anti-realist or error theoretic arguments that reject the objectivity and categoricity of moral norms, I am not necessarily concerned only with such arguments. While many philosophers who take the empirical approach to ethics do reject the categoricity of moral norms, it is not an essential aspect of the position. An empirical ethicist could conceivably think that there are objective moral facts that can be

empirically discovered. The defense of companions in guilt arguments that follows is not intended to defend the traditional philosopher's objective understanding of norms against the empirical philosopher's interpretation of them as subjective. My aim is just to show the close similarities between ethics and epistemology with regard to questions like these in order to motivate looking at how epistemologists have attempted to answer the sorts of challenges this project has been concerned with and consider how they can be adapted and used by ethicists.

Intuitively, companions in guilt arguments are appealing. Moral error theory has many forms, all with slightly different arguments, motivations, and evidence, from Mackie's Argument from Queerness (Mackie 1977), to the evolutionary error theory defended by Street (2006).<sup>84</sup> This makes error theory difficult to respond to, because "error theory" does not refer just to one argument or one set of arguments. By showing that the arguments that establish various moral error theories overreach, however, and lead to denying the existence of epistemic norms – which can be shown that we should accept for other reasons – the CIGAr establishes that something must be wrong with the arguments for moral error theory, regardless of the form they take. Of course, this only works if both the parity premise and epistemic existence premise are actually true. In this section, I look at objections to the CIGAr, focusing particularly on Cowie's objections, and argue that they are not sufficient to undermine the effectiveness of the CIGAr.

Cowie has two main arguments against the CIGAr: an argument that the parity and epistemic existence premises are in tension with each other (2014), and a "master argument" (2016). In the latter paper, Cowie claims that the former argument was "imperfectly developed" and is picked up on and developed further in his master argument (2016, 108). However, the two sets of arguments are distinct enough that they are worth considering separately.

In his first paper on the topic, Cowie argues that arguments for the epistemic existence premise either undermine the parity premise – his "objection from disparity"

---

<sup>84</sup> Street's argument is discussed in chapter 1.

(410) – or they “render the companions in guilt argument dialectically redundant” (ibid.). That is to say, accepting the epistemic existence premise in the first place requires rejecting moral error theories on other grounds, and thus the CIGAr is unnecessary. Since the epistemic existence premise is necessary to the success of these arguments, Cowie concludes that any version of the CIGAr is necessarily self-undermining, and so will not work. For our purposes, the redundancy argument is not particularly important, and so it will be set aside.

The arguments Cowie is concerned with are the various forms of the argument that epistemic error theory is self-defeating. As Cowie summarizes them, these arguments are relatively simple, stemming from the fact that a rejection of the existence of any categorical epistemic norms also entails a rejection of the existence of any categorical reason to accept epistemic error theory. Trivially, either epistemic error theory is true, or it is false. If it is false, then we should not believe it. If it is true, however, then there does not exist any reason to believe it, and so we still should not believe it. Thus, the argument goes, epistemic error theory is self-defeating (Cowie 2014, 411).

Cowie notes, however, that while this argument does seem to potentially undermine epistemic error theory, an analogous argument does not show moral error theory to be self-defeating (ibid.). This may not seem like a particularly big problem; if there were straightforward arguments showing that moral error theory to be self-defeating or otherwise inconsistent, it would be unnecessary to turn to the CIGAr to defend moral facts. But such arguments hinge on the idea that moral norms and epistemic norms share similar relevant features, and Cowie argues that this difference is relevant enough to create an important disanalogy between the two types of norms, and thus to undermine the parity premise. Specifically, he argues that the core point of error theorists’ arguments is to show that the property of categoricity is metaphysically and epistemologically problematic. And while the argument from self-defeat shows that it is inconsistent to reject the existence of categorical epistemic norms, it does *not* show (nor does it attempt to show) “that epistemic facts are otherwise metaphysically and epistemologically



unproblematic” (Cowie 2014, 412). Instead, Cowie argues, it shows that epistemic norms have some “special property” that means we should reject epistemic error theory (ibid.).<sup>85</sup> But since this does nothing to make categoricity less problematic, it does not necessarily give us any reason to accept the existence of anything else that is categorical, but lacks the “special property”. So, Cowie argues, since moral norms do not have this special property, there is a relevant disanalogy between epistemic and moral norms, and so the parity premise is undermined.

There are a number of ways to push back against this type of argument. Rowland, for instance, argues that the disparity disappears if moral and epistemic norms are both instances of the same type of relation (evidential support relations, for instance), and further that they should be understood as the same type of relation, and so Cowie’s objection does not work (Rowland 2016). While I think that Rowland’s argument poses significant problems for Cowie’s rejection of the CIGAr, both it and Cowie’s initial argument place too much importance on the extent to which epistemic and moral norms need to mirror each other for the parity premise – and thus the CIGAr – to succeed. They do not need to be perfectly analogous to each other, nor does the CIGAr claim that they are. The CIGAr claims that a) moral error theorists reject moral facts on the grounds that they are categorically normative and b) epistemic norms are also categorically normative, but clearly exist. In other words, the CIGAr is meant to show that error theoretic argument overreach and deny the existence of a category that undeniably has members. To succeed, then, it only needs to show that *something* falls into that category that error theorists reject.

To see this more clearly, it helps to note that the CIGAr is not a positive argument, seeking to establish the existence of moral norms. Rather, it is a purely negative argument, seeking only to establish that the arguments of error theorists do not succeed in proving their conclusion. Since this is the case, we should consider the aim of the CIGAr to be to – somehow or other – undermine some of the support for error

---

<sup>85</sup> Specifically, that accepting epistemic error theory is self-defeating

theoretic arguments. The general form of the argument in this case is as follows: 1) Categorical normativity is metaphysically and/or epistemically problematic, to the extent that we should reject the existence of things that purport to categorical normativity.<sup>86</sup> 2) Moral norms are/purport to be categorically normative. 3) Therefore, we should reject (or at least be deeply skeptical of) moral normativity. The CIGAr works by pushing back against that first general premise. The initial premise commits the error theorist to the claim that nothing exists that has the property of categorical normativity. The truth conditions of this premise are fairly clear: it is true if and only if nothing has the property of being categorically normative, and so the CIGAr aims to show that something has this property, and thus that the first premise is mistaken. The candidate in this case is epistemic norms. They are categorically normative, but error theorists should be generally unwilling to deny their existence.

Importantly, any other properties possessed by epistemic norms are unimportant to this discussion. The error theorist has claimed that there does not exist anything with property *P*, and the companions in guilt argument responds by pointing out that there does indeed exist some *x* that has property *P*. Responding that there is some special property that *x* has in addition to *P* is irrelevant.

Consider an analogous argument: Gina, a unicorn realist, is debating Jake, an equine error theorist, about the existence or non-existence of unicorns. Jake argues that unicorns don't exist, because there is no such thing as a four-legged equine, and so we should abandon any belief in animals that have that property. Gina, wisely, responds that horses are four-legged equines, and not only do horses exist, but Jake has seen and interacted with them repeatedly.<sup>87</sup> At this point, the correct thing to take from this is that – even if Jake is right, and there are no unicorns – his argument for this point must fail. There may be other arguments that will successfully establish that unicorns do not exist,

---

<sup>86</sup> This premise can be established through various arguments, including the argument from queerness, or evolutionary debunking arguments, among others.

<sup>87</sup> Call this the “companions in hooves” argument

but this one will not. Further, pointing to a “special property” possessed by horses, but not by other four-legged equines (because “four-legged equine” remains a metaphysically/epistemologically problematic concept) does not redeem this argument. There may well be some property possessed by unicorns in virtue of which we should doubt their existence, but it cannot be that they are four-legged equines. Similarly, if there is some property possessed by moral norms in virtue of which we should be skeptical of their existence, it must be something other than their categorical normativity.

Cowie’s “master argument” attempts to bring together the concerns of various arguments used to discredit companions in guilt strategies. Cowie sums it up as a dilemma: “Either categorically normative epistemic reasons for belief are evidential support relations or they are not. If they are, then the parity premise is false. If they are not, then the epistemic existence premise isn’t established. In either case, the companions in guilt strategy fails.” (Cowie 2016, 127). As responses that reject the existence of epistemic norms will be addressed later, we will put that horn of the dilemma aside for now. The first horn runs into the same problem that we have already seen in his other argument: It is not necessary that moral and epistemic norms be alike in all relevant ways for the CIGAR to succeed. What is relevant is that they are alike in the ways that are relevant to the error theoretic argument: their categoricity. Cowie argues that understanding epistemic reasons for belief as evidential support relations reduces them to something that is not categorically normative, (Cowie 2016 123-4). But saying that something being a reason for believing *x* just means that it raises the probability of *x* being true, we have only pushed the normativity back a step. We can just as easily understand the claim, “*x* raises the probability of *y*” as meaning, at least in part, “*x* gives you more reason to believe *y*” which is still a normative claim as it still action-guiding, concerned with what one ought to do; in this case, increase one’s confidence in *y*.

The defense of the CIGAR I have been developing bears some similarities to another argument that has been deployed by Thomas Nagel, Richard Dworkin, and others, and it is worth taking a moment now to look at that argument and how the CIGAR differs. Known as the Moorean argument, its basic form is as follows:

1. If moral error theory is correct, then some clearly morally wrong thing (throwing babies onto bonfires for fun, torturing small children for fun, etc.) would not be morally wrong.
2. It seems to me to be more certain that the clearly morally wrong thing is morally wrong than that any of the premises supporting moral error theory are true. So,
3. Moral error theory is false. OR We should believe in the moral wrongness of the clearly morally wrong thing and reject error theory.

The basic strategy is fairly clear and a fairly common philosophical approach. We are presented with a dilemma: the arguments for moral error theory are incompatible with continued belief in certain moral facts. Faced with this dilemma, we have to choose what to reject: moral facts, or the premises which support moral error theory. A decision metric is suggested – specifically, hold onto the belief or set of beliefs in which one has the greater confidence – and based on that, the conclusion is reached. Certain moral facts appear highly likely to be true, and our degree of belief in those moral facts is (or ought to be) higher than our degree of belief in the premises that support moral error theory. And so, faced with this dilemma, moral error theory is the one to reject.

This argument should look familiar, as it is essentially the same argument as the one Nagel deploys in *Mind and Cosmos*, which was discussed in the first chapter of this work. Here, as in that case, this argument leaves much to be desired. In the case of his argument from *Mind and Cosmos*, I argued that it is a mistake to have a higher degree of belief in moral facts than in the theories which support the entire field of biology. Similarly, in this case, it is fairly easy to give an account of why holding a higher degree of belief in moral facts than in the arguments that support moral error theory is mistaken. Jonas Olson, for instance, points out that evolutionary debunking arguments provide an explanation of why we would not only believe certain moral claims regardless of their truth value, but even have a very high level of confidence in their truth (Olson 2011).

This is also part of what Street's Darwinian Dilemma<sup>88</sup> shows: even if all our moral beliefs were false, we would still believe them to be true. More generally, error theoretic arguments often provide an explanation for why we would strongly hold our moral views even if there is not any actual reason to hold them. Once such a defeater exists, simply citing strong confidence in the truth of moral claims is no longer a sufficient reason for holding onto our belief in moral facts over the premises of error theoretic arguments. Rather, citing that confidence in moral facts – absent a solid argument for why that defeater is not enough to undermine confidence – amounts to little more than refusing to update beliefs in accord with new evidence.

By contrast, the CIGAr does not rest on subjective claims about degrees in confidence in the premises of error theoretic arguments or moral facts. Instead, all it relies on is the tension that exists within error theoretic arguments between the denial of moral normativity and an unwillingness or inability to reject epistemic normativity. The existence of epistemic norms is important to the success of the CIGAr, but it is not the level of confidence we have in epistemic norms that makes the argument work. Instead, it is the fact that denying epistemic norms is either self-contradictory, or at least undermines any reason we would have to believe error theoretic arguments that the CIGAr its force. Thus, the response from the error theorist that her arguments already show why we would mistakenly continue to believe in moral facts even if they were false do not apply in this case, as that is not relevant to the CIGAr.

Rowland (2016, 165) suggests quantitative parsimony as a potential route for opponents of the CIGAr to take, but as he points out, this is ultimately a weak line of argument. According to the quantitative parsimony requirement, we should not posit more members of a given class than is necessary (*ibid.*).<sup>89</sup> By this argument, although we have reason to admit that categorically normative reasons exist in the form of epistemic

---

<sup>88</sup> Discussed in detail in chapter 1.

<sup>89</sup> In contrast to the typical (or qualitative) understanding of parsimony, according to which there is a prohibition against unnecessarily expanding the number of classes or categories.

reasons, we should nevertheless not accept the existence of any other sort of categorically normative reason for the sake of keeping our theories metaphysically clean. This appears to be in keeping with Cowie's arguments against companions in guilt strategies. The thrust of his arguments is that even if we accept the existence of categorical epistemic norms, we are not justified in accepting the existence of moral norms. One potential route he can take to support this argument is an appeal to quantitative parsimony: even if there are categorically normative reasons, we should not posit any more members of that category than is absolutely necessary, and so we should not accept the existence of categorically moral norms.

There are two significant problems with this account. First, it does not track restrictions we actually place on our theories. Once we have established that epistemic norms exist, for example, it is not a particular virtue of a theory if it claims that there is a smaller number of those norms than other competing theories would hold. If that theory can show that other theories posit too great a number of epistemic norms because they include ones which are redundant, that may count in favour of the theory, but it is not the mere fact of that theory's quantitative parsimony that privileges it over competitors. While quantitative parsimony does count positively for a theory, it can be outweighed by any number of factors, including that a newly-positing entity helps fill out an explanation. As Rowland points out, the fact that positing the existence of categorical moral norms helps make sense of much of our moral discourse – which even most error theorists will admit – means that positing their existence helps provide an explanation for a phenomenon, and thus is enough to outweigh concerns of quantitative parsimony (2016, 165-6).

Second, such an objection to the CIGAr misunderstands what the aim of the argument is. It amounts to arguing that, although denying epistemic norms exist is self-defeating and denying categorical normativity entails denying the existence of epistemic norms, the CIGAr has nevertheless not proven that specifically moral categorical norms exist. This is true, but it is also irrelevant. As stated above, the CIGAr is a debunking argument. It succeeds by showing that the error theorist's argument fails to establish its conclusion, not by providing proof that categorical moral norms exist. So, we should not

expect the CIGAr to provide specific evidence for the existence of moral norms. Rather, it succeeds so long as it shows that error theoretic arguments overreach. Put differently, what is at issue in this discussion is not – directly – whether or not there is positive evidence for moral norms, but rather whether or not error theorists' arguments force them to overcommit and deny the existence of epistemic norms, and whether or not there is a relevant analogy between epistemic and moral norms. As long as this is established and the error theorist is unwilling to deny the existence of epistemic norms, companions in guilt arguments are able to succeed.

## 4.2 Limits of Companions in Guilt Arguments

There are, however, significant limitations to the reach of the CIGAr. As we have seen, it does not provide a positive argument for the existence of categorical norms, focusing instead on debunking skeptical arguments. And so, while this argument is useful for exposing the flaws in certain types of other arguments, it cannot by itself provide support for the existence of moral norms. Furthermore, not all morally skeptical positions draw their skepticism from characteristics that moral norms share with epistemic norms. If the skepticism is motivated by some feature particular to moral norms, then the parity premise fails, as does the overall argument. Finally, the CIGAr only succeeds if the skeptic is unwilling to deny the existence of categorical epistemic norms. If the moral skeptic is willing to bite that bullet, then the CIGAr has no force. The first of these limitations has already been addressed, and is more a problem with what people expect from the CIGAr, as opposed to a problem with the argument itself. The other two, however, are potentially more significant issues, and are worth examining in more detail with a particular focus on how they impact the arguments we have considered up to this point.

The question to ask for our purposes is, does the CIGAr apply to the objections to traditional approaches to ethics raised by empirically-minded philosophers? After all, we are not concerned with all possible error theoretic arguments. All that matters for our purposes are the arguments that motivate or arise from the empirical approach. First, let us consider Street's Darwinian Dilemma. As discussed previously, this argument turns on the idea that our moral intuitions are shaped by evolutionary pressures which select for

what is evolutionarily adaptive rather than for any coherence with moral facts. In this particular case, we can – at least *prima facie* – apply this argument to epistemic judgments as well. If, as Street claims, our fundamental tools for forming judgments and intuitions are coloured by our evolutionary history, then this is true of not only our moral judgments and intuitions, but also our epistemic ones. Indeed, Street herself has argued for exactly this position, claiming that, “antirealism about practical reasons hasn’t received a full defence until antirealism about epistemic reasons has been defended as well” (Street 2013, 213-4) and arguing for this epistemic antirealism along similar evolutionary lines. Although Street puts it differently, this amounts to an endorsement of the core of the CIGAr: skepticism about moral norms entails skepticism about epistemic norms. It is, of course, noteworthy that Street’s solution is to reject epistemic realism along with moral realism, but we will return to this point later. For now, what is important to bear in mind is that this sort of evolutionary debunking argument against moral norms also provides an argument against epistemic norms.

Similarly, the various cognitive biases considered earlier frequently pose every bit the challenge to epistemic norms that they do for moral norms. Some, like in-group bias,<sup>90</sup> have a bigger impact on moral judgments than epistemic ones, as they affect our understanding of how we relate to others and what behaviours are called for, not judgments about what is true, or what counts as evidence.<sup>91</sup> Sinnott-Armstrong’s master argument, previously discussed in chapter 1, captures the problem that cognitive biases form for the reliability of moral intuitions. As a reminder, the master argument proceeds as follows:

- (1) If our moral intuitions are formed in circumstances where they are unreliable, and if we ought to know this, then our moral intuitions are not justified without inferential confirmation.

---

<sup>90</sup> Our tendency to like people/things that we see as being similar to us in some relevant way.

<sup>91</sup> These moral judgments are not entirely separable from epistemic concerns, but questions of epistemic injustice are outside the scope of this project.



- (2) If moral intuitions are subject to framing effects, then they are not reliable in those circumstances.
- (3) Moral intuitions are subject to framing effects in many circumstances.
- (4) We ought to know (3).
- (5) Therefore, our moral intuitions in those circumstances are not justified without inferential confirmation. (Sinnott-Armstrong 2008, 52)

Note that nothing in this is unique to moral intuitions, and the argument proceeds just as well with the word “moral” replaced with the word “epistemic.” And so, in this case as well, the companions in guilt argument applies unless philosophers appealing to cognitive biases and framing effects are willing to also reject the existence or reliability of epistemic intuitions and norms.

Similarly, the motivations that Prinz and Nichols present for appealing to folk intuitions and rejecting the expertise of moral philosophers also apply equally to epistemology. As discussed earlier, Nichols suggests that, “when the philosopher-experts disagree on some central feature of folk concepts, it seems particularly useful to see what the folk have to say” (2008, 399), while Prinz argues that folk intuitions are potentially free of the biases of professional philosophers (Prinz 2007b, 274). The motivation in its most basic form is that ethicists have yet to “solve” the problems with which we are concerned, and instead of continuing to let ethicists debate each other endlessly, we can turn to folk intuitions to help break the deadlock, or push the argument forward. At this point, however, it should be clear that epistemology is no more “solved” a field than ethics, and epistemological problems are no less heavily debated than moral ones. We might suggest that non-philosophers are less concerned with the problems faced by epistemologists than ethicists, claiming that folk intuitions about what is right and wrong are clearer and more useful than ones about preface or lottery paradoxes, for instance, but this is not relevant. The suggestion from Nichols and Prinz is that philosophers have stalled or are otherwise unable to move their arguments forward, and so we should look to folk intuitions for clarity. To the extent that this is true of moral philosophy, it is also true of epistemology, and so the companions in guilt argument would apply in this case as well.

So, it appears that the CIGAr presents at least some challenge to the major empirical objections to traditional approaches to ethics. The question now is, what should we conclude from this? It is tempting to overreach and claim that this shows these arguments to be fundamentally flawed or too broad such that we should reject them. That would be too hasty. Thus far, all that has been shown is that these arguments satisfy the parity premise of companions in guilt-style arguments. Here, however, a problem that I alluded to when discussing Street's argument above becomes particularly relevant: the parity premise is only one of the two premises of the CIGAr, and the limitations of the epistemic existence premise also impose important limitations on this argument.

Up to this point, I have been focusing on the parity premise and taking the epistemic existence premise for granted, but as the overall argument is only as strong as strong as this premise. I said previously that the defense of this is often more straightforward, arguing that denying the existence of epistemic norms is incoherent or self-defeating. At the very least, it is a dialectical struggle. Arguing that epistemic reasons for belief do not exist necessarily requires providing reasons that one's interlocutors should accept the argument, while simultaneously asserting that those arguments do not actually provide categorical reasons to believe the argument.

Despite the obstacles such an argument presents, it should come as no surprise that some philosophers do indeed reject the epistemic existence premise, and have argued that denying the existence of categorical reasons for belief is not self-defeating. Stan Husi (2013), for instance, provides exactly such an argument, showing that reasons skepticism can be supported on the basis of non-authoritative (or non-objective) norms. Similarly, as mentioned above, Street (2013) defends epistemic anti-realism. These are only two among many examples of arguments rejecting the objective authority of epistemic norms. Indeed, it is not hard to imagine our epistemologist colleagues being both surprised and delighted to find out that their ethicist counterparts have such confidence in them as to consider epistemic norms beyond reproach. Here we see the more crucial limitation of the CIGAr: it hinges on the notion that the authority of epistemic norms is unassailable, but this is not so widely accepted. As we determined before, the success of the CIGAr means that moral skeptics cannot reject the existence of moral norms without also rejecting

epistemic norms. Unfortunately, however, this is not so difficult a bullet to bite for anti-realists.

Patrick Clipsham (forthcoming) has developed an alternate approach to the CIGAr which avoids this response from error theorists. Rather than relying on the existence of epistemic norms, he points out that, “one cannot meaningfully participate in academic philosophical dialogue without tacitly accepting the categorical, objective, or authoritative correctness of certain norms of philosophical argumentation” (4). These norms include ones like the principle of charity or ontological parsimony. Clipsham argues that there is a tension arising from the error theorist accepting the categoricity of these metaphilosophical norms while rejecting the categoricity of moral norms. Structurally, this argument is fairly similar to the “typical” companions in guilt strategy: Clipsham establishes that it is very likely that error theorists accept the categoricity of metaphilosophical norms of argumentation – analogous to the epistemic existence premise – then argues that this means they are not justified in rejecting moral norms.

The central point of Clipsham’s argument is his claim that it is very likely that error theorists accept the categoricity of metaphilosophical norms. As he points out, the process of engaging in the sort of debates and arguments that philosophers – including error theorists – engage in requires holding to certain norms of argumentation. Further, their behaviours cannot be properly explained as just a hypothetical commitment to these norms. The error theorist would not be satisfied, for instance, by an opponent insisting that it was fine that she responded to an uncharitable understanding of the error theorist’s position because the principle of charity has no authoritative force, and, as it happens, she does not particularly care about that norm (10). The maximally consistent error theorist, in this situation, would have to “concede that their opponent has not committed any kind of objective philosophical error, but rather is merely violating a contingent, subjective, or hypothetical norm that is in no way authoritative, and that no one has any particular reason to accept” (ibid.). But no philosopher would respond to such a claim by admitting that their interlocutor has a point, and although they subscribe to a norm by which one should interpret one’s opponents charitably, they cannot fault their opponent for not doing so. It is, quite simply, not how we interact with each other.

Although I think this argument succeeds in its aim of demonstrating a tension between the error theorist's tacit endorsement of metaphilosophical norms and his rejection of moral norms, it does not quite serve our purposes here. As Clipsham points out, "the metaphilosophical companions in guilt argument does not show, or claim to show, that moral realism is true [. . .] or even that error theory is false [. . .]. Rather, it seeks to demonstrate the plausibility of attributing an instability to any attempt to defend the error theory from within academic philosophy" (14). As the aim of this project, however, is to provide a framework to defend moral philosophy against empirically-grounded error theoretic arguments, Clipsham's argument does not go far enough. So, while the metaphilosophical CIGAr gets at a real tension between the behaviours and professed beliefs of the error theorist, we need to push the argument further than this.

Where, then, does that leave us? Is the CIGAr ultimately irrelevant, unable to defend moral realism in the face of epistemic skepticism, making this exploration of the argument a waste of time? This may be a tempting conclusion to draw, but it is also incorrect. There are two important things to take from the discussion so far. First, the various arguments used to motivate the rejection of the traditional approach by proponents of the empirical approach also entail a similar rejection of epistemic norms. Further, as I have said, the purpose of the companions in guilt argument is not to provide positive proof for the existence of categorical moral norms. What it does successfully demonstrate is that epistemic norms and moral norms stand or fall together. That is to say, moral skepticism entails epistemic skepticism. This leads us to the second important thing to note: Realists and skeptics can agree that moral norms and epistemic norms are "companions in guilt." As we have seen, arguments that successfully undermine this argument do not argue that moral and epistemic norms are relevantly different; they argue instead that categorical epistemic norms should also be discarded. Taking these two things together leads us to a potentially useful route for our inquiries: if epistemic and moral norms are companions in guilt, they can also be companions in defense, as I argue in the next chapter.

## Chapter 5

### 5 Empirically-Informed Moral Philosophy

As we saw in the previous chapter, rejecting moral norms in virtue of their fundamental normativity entails a rejection of epistemic norms. The companions in guilt argument presents this as a reason to reject such error theoretic arguments, as a rejection of epistemic normativity is purportedly self-defeating. Despite this, arguments exist that appear to show that epistemic norms are vulnerable to the same sort of debunking to which moral norms have been subject. If this is the case, then the companions in guilt argument is irrelevant. Moral and epistemic norms are both "guilty" and the argument does nothing to rescue moral normativity. But this also leads us to a potential way to solve the problems we have been facing.

In this chapter, I examine the potential for epistemologists and ethicists to act as what I am calling, "companions in defense." That is to say, the potential for arguments used by epistemologists as responses to epistemic debunking arguments to be adapted by moral philosophers as responses to the sorts of challenges I have pointed out in previous chapters. In the first section, I look at how epistemologists have responded to the sorts of empirical challenges that have posed problems for moral philosophy and consider how moral philosophers could employ similar arguments. Sections two through five return to the role of intuitions in moral philosophy and explore how to make use of our moral intuitions in a way that is compatible with the methodological restrictions laid out in the previous chapters. In the final two sections, I bring all of this together, and present a preliminary version of an empirically-informed moral philosophy that can present the sort of prescriptive normativity we want from our account of ethics while still being adequately responsive to empirical and scientific concerns. I conclude that information from cognitive science and evolutionary biology need not pose a threat to our ability to conduct philosophical inquiry into moral questions.

## 5.1 Companions in Defense

Epistemology is a perfect place to turn to look for arguments that can potentially be analogously applied to ethics. In addition to being the source of normativity pointed to by companions in guilt arguments to help rescue moral normativity, the two areas also have many similarities. Like ethics, epistemology is profoundly norm-guided. In addition to this – as I argued in the second chapter – moral philosophy is at its core a prescriptive task, seeking to find the norms we should hold, not (necessarily) the ones we actually do. This is every bit as true of epistemology: the aim is not to describe how people actually come to form beliefs. Most Bayesians, for instance, do not think that people actually have the probabilistic relations between all their beliefs in mind and update their priors in appropriate proportion to any new evidence. Rather, they think this is an idealization, reflecting how a perfectly epistemically rational agent would adjust and weight their beliefs. Finally, as I gestured at above, epistemology faces similar challenges to its traditional approaches from evolutionary and neuroscientific perspectives to those faced by moral philosophy. Combining this with the fact that companions in guilt arguments rely heavily on the relative similarity of moral and epistemic norms means that looking to ways epistemologists have addressed these issues is informative to this project.

There are some particular advantages to this approach. While epistemology is certainly not a “solved” area of philosophy, it is somewhat easier to judge if a proposed epistemic norm is “good” than it is to similarly judge moral norms. That is to say, we have some method of testing whether or not an epistemic norm is a good one. If the norm leads to a series of true beliefs and is generally good at avoiding false ones, it’s a good norm. There are, of course, issues of skepticism, and questions of exactly what it means for a proposition or belief to be true, and how we can know that something is true. Despite these concerns, however, we can generally know, for instance, that observations made through a telescope more reliably reflect reality than those made through a kaleidoscope. We do not have quite so clear a way of evaluating whether or not any purported moral norms are comparably correct. We know, of course, that an ideal norm will lead to an agent doing what is right if she follows it, but we can’t observe whether or

not a moral norm is leading us to correct, novel predictions in the same way we can with an epistemic one.

Similarly, we also have a better sense of what the “goal” of our epistemic norms is. We can, in a sense, claim that there is only one central epistemic norm: believe all and only true sentences. Any other epistemic norms are ultimately in service of this; they help us to more reliably get to the truth and avoid error. Radical skeptics accomplish the latter part of this goal by believing nothing and thus not believing anything false, while the overly credulous person can accomplish the former part by believing every possible proposition. The overall aim is to try to find the right balance between these two extremes so that we can avoid error like a skeptic, but believe true statements like a credulous person.<sup>92</sup> Again, this goal is not as clear in the case of ethics. We have a central aim: something like being good, or knowing and acting in accord with what is right, but what that means is exactly what is at issue in our debates. As a result, it’s much harder to clearly pin down what counts as moral progress, or know if we are succeeding in moving our understanding forward.<sup>93</sup> We could, as Campbell and Kumar (2013) do, take a subjectivist route and suggest that moral progress consists of increasing our moral consistency, and thus look to see if a proposed norm causes our overall moral picture to be more coherent, but this is an unsatisfactory response. Trivially, there are two ways to respond to finding out that belief *A* and belief *B* are in conflict with one another: reject/revise *A*, or reject/revise *B*. For example, if – as in Singer’s famous example (Singer 1972a, 232-3) – a person finds that there is a tension between his willingness to wade into a pond to save a drowning child and his unwillingness to donate to help with famine relief in a distant country, he can resolve this either by deciding that he ought to donate to famine relief, or by deciding that he shouldn’t save the drowning child. Either

---

<sup>92</sup> This is an oversimplification of the epistemological project, but it serves well enough for our purposes here.

<sup>93</sup> Recall that this complaint was one of the cited reasons for adopting the empirical approach we considered in chapter 2.

one of these options makes his overall moral position more consistent,<sup>94</sup> but it is also a highly unsatisfying response. Such a response treats the question, “I think that I ought to do  $x$ , but is that really right?” as being incoherent, since the “correct” answer is just whatever increases consistency. As I argued in the second chapter, however, this sort of answer does not answer the sorts of questions we are looking at when we engage in moral philosophy.

Alternatively, we could take the objectivist route and argue that a norm makes moral progress when it brings our moral views more in line with objective moral truth, but what this means is exactly the question that is at the core of moral philosophy, and the subject of the most disagreement. Kantians would argue that norms that better match up with objective moral duty accomplish this goal, while classical utilitarians would argue that we should look to norms that increase overall happiness. We have tools for addressing these debates, but again, this is exactly what is at issue in moral philosophy. So, this response only pushes the question back a step.

Considering all of this, it seems helpful to look at how epistemologists have answered the sorts of questions that have plagued us throughout this project. More specifically, it seems useful to briefly look at how epistemologists have dealt with empirical challenges to traditional approaches to epistemology, what this reveals about the methodological assumptions being made, and if and how we can use this to inform our attempts to bridge the gap between the traditional and empirical approaches to ethics, and to resolve the problems both of them face. To that end, a few caveats are appropriate. First, a thorough survey of the state of contemporary epistemological debates (even limited purely to empirical challenges to traditional approaches) is well beyond the scope of this project. Instead, I will be focusing on one approach that is particularly informative, namely the approach taken by Kaija Mortensen Jennifer Nagel. Second, it is certainly not my intention to claim that any solutions epistemologists may have found to the sorts of problems I have been focusing on can be immediately applied to ethics, and

---

<sup>94</sup> Assuming that either choice meets the other criteria Campbell and Kumar lay out, but we can assume that for these purposes.



thereby solve everything. Although ethics and epistemology are closely related in virtue of being areas with a particular focus on normativity, and although similar empirical challenges have been raised in both areas, they are nevertheless different enough as to require unique solutions and approaches to their problems. Rather, my hope is that work already done in epistemology can provide insight to the general direction of an approach that we can take. With that in mind, let's now turn to epistemological responses to empirical challenges.

Mortensen and Nagel (2016) identify three challenges posed by experimentalists, which they refer to as the *Diversity Challenge*, the *Ignorance of Folk Thinking Challenge*, and the *Questionable Evidence Challenge* (54-5). According to the Diversity Challenge, traditional philosophical approaches are insufficiently sensitive to the variations in intuitions among cultural and other demographic lines (54). The Ignorance of Folk Thinking Challenge holds that philosophers “make assumptions about how people ordinarily think” (55) but are often incorrect, and people do not in fact think about morality, knowledge, or other philosophical topics in the way philosophers think they do (ibid.). The final challenge – the Questionable Evidence Challenge – claims that, even if the above two problems did not exist, “philosophically significant intuitions may be inappropriately sensitive to such apparently irrelevant considerations as the order in which the cases are presented” (ibid.).

These challenges should seem familiar, as they are exactly the challenges to traditional approaches to ethics discussed in the first two chapters. The Diversity and Ignorance of Folk Thinking challenges are explicitly what motivate Prinz and Nichols to seek the opinions of non-philosophers and implicitly reject the idea that philosophers can be moral experts.<sup>95</sup> Similarly, the Questionable Evidence Challenge is what is at issue when philosophers express concerns about cognitive biases, what leads to Sinnott-Armstrong's “master argument” (Sinnott-Armstrong 2008, 51-2), and is what motivates Street's Darwinian Dilemma (2006). These are the major issues that motivate skepticism

---

<sup>95</sup> Cf. Prinz 2007b, 274-9; 2009, 134-8; Nichols 2008b, 397-9; and section 2 of chapter 2 of this work.

about the use of intuitions in moral philosophy, which is the most problematic objection to the traditional approach. And so, if we can find a solid answer to these challenges from epistemology, we are potentially that much closer to finding a way to handle the challenges we face.

Mortensen and Nagel provide a response to each of these three challenges, all working along a similar line. They point out that each of the three challenges fails to actually be supported by further/more detailed empirical investigation. Various studies have failed to show any significant variance in intuitions about epistemological scenarios along cultural, gender, age, or socioeconomic lines (Mortensen and Nagel, 58-60), seeming to show that the Diversity Challenge is unfounded. Similarly, with regards to the Ignorance of Folk Intuitions Challenge, the studies cited by Mortensen and Nagel indicate both that philosophers are quite good at predicting the intuitions of non-philosophers, and that there is not a dramatic difference between the intuitions of the two groups (60-1).

The response to the Questionable Evidence Challenge is somewhat more complicated. First, as Mortensen and Nagel point out, “[w]anting sources of evidence to be sensitive only to the truth or falsity of what is relevant is wanting something very strong indeed” (62), as this is something that is not even true of sensory perception. The relevant types of error are ones where we’re not necessarily aware of the ways in which our intuitions can go wrong, or where we cannot predict or control for the error. However, as they further point out, experimentalists, “are overestimating the seriousness of the variation here, or [...] they are underestimating the capacity of armchair philosophers to spot problems from the armchair” (ibid.). That is to say, looking at further empirical data suggests that framing effects and other similar issues are not nearly as strong as the experimentalists suggest, and philosophers are generally good at being aware of when our intuitions are generally less reliable, or need to be treated with more caution. As an alternate response, they also suggest as a possible line of argument “that philosophers’ answers are on a firmer foundation than those of laypeople” (ibid.), which is to say – as I argued in chapter 2 – that philosophers should be understood as having special expertise in virtue of our work and training.

Initially, this may not seem like the most helpful set of responses for our purposes. As I have already argued, the empirical approach to philosophy rests on a series of methodological assumptions that cannot be supported, yet these responses rely on turning to yet more empirical data. If this response is essentially just repeating the empiricist ideas I have already criticised, it would add nothing to our conversation. As has been the case throughout this project, however, the particular details of the argument are less important than what those arguments tell us about the methodology being employed. In this case, Mortensen and Nagel are defending an “armchair-friendly” understanding of experimental philosophy, examining “ways in which experimental methods can complement and even strengthen armchair-style philosophy” (53). And so, the studies they cite are being cited not because the empirical data are all there is to epistemological inquiry, but because the data can help to make the work done from the traditional perspective simpler, or to clarify things. They point to several case studies of philosophers using experimental data not to undermine traditional approaches, but to help push debates forward.

Their analysis of one particular case study is especially informative. They cite an ongoing problem in metaphysics, and specifically mereology: metaphysicians want their theories to match up with folk metaphysical views as much as possible, but often have difficulty accomplishing this (65). In response to this problem, David Rose and Jonathan Schaffer (2017) use empirical data to show that the folk understanding of mereology is teleological, then argue that teleological accounts are necessarily incorrect, “therefore giving us reasons to believe that folk mereology need not act as a constraint on our philosophical theories of mereology” (Mortensen and Nagel, 65). Here, as they analyze it, the empirical testing of folk intuitions does not serve as a means of solving the philosophical question at issue, nor even really as evidence to support some philosophical thesis. Instead, it is being used to help resolve an existing methodological challenge and push a debate forward. The primary work being done in the overall argument is the traditional philosophical argument that teleological accounts of mereology are false. The empirical study only plays a small supporting role.

This sort of approach is common in Nagel's work. She has regularly been concerned with defending the evidential value of epistemic intuitions and epistemologists' uses of thought experiments and other case studies in the face of experimental debunking arguments (Nagel 2012; 2013; Boyd and Nagel 2014). These defenses are accomplished through a hybrid approach. On the one hand, Nagel often employs an empirical approach to show that the studies cited by experimental philosophers are not as robust as they would need to be to support their claims. At the same time, she also employs a traditional approach, arguing "from the armchair" that the studies in question – even if they were robust and successfully replicated – would not show what they purport to show.

Nagel draws an analogy to perception. The upshot of experimental work in epistemology is supposed to be "its promise to give us something more than generic skepticism, its claim to deliver an empirically supported challenge to a default position of trusting epistemic intuition" (Nagel 2013, 180). That is to say, it motivates skepticism about the use of intuitions, but is not just a rehash of standard skeptical arguments that entail that we also cannot rely on perception. However, it is also true that our perception is not perfectly reliable. So, for this empirical project to succeed, it must not only show that epistemic intuitions are prone to error, but also that they are prone to error in a way that perception is not. One of Nagel's arguments is that the empirical arguments fail to do this.<sup>96</sup>

Taking all of this together, we are in a good position to explore the methodology underlying this epistemological response to empirical challenges with a particular eye to the issues considered in the first two chapters. There are two questions to answer here. First, how does what we have seen here fit with the methodological concerns that we

---

<sup>96</sup> This general point is also relevant to the empirical approach to ethics. While this approach displays skepticism towards the traditional methods of ethics, it is not a generally skeptical position. That is, it does not hold that knowledge in general is unreliable, or that we cannot trust our perceptions. In fact, any sort of overall or radical skepticism is deeply incompatible with the empirical project. At the very least, proponents of this approach must think that scientific inquiry is a reliable means of producing knowledge. Otherwise, they would also have to hold that the scientific data they cite is no more useful as information than the intuitions they reject.

have already raised? Second – and more importantly – what does this add to our understanding of methodological approaches to ethics? To the first question, if this approach commits us to the wrong position on the methodological questions posed earlier, then it cannot help us to resolve the tension between the two approaches to ethics. Similarly, it also cannot be of any use if it does not add anything new to our approach.

To answer the first question, we need to consider the methodological challenges raised in the first two chapters. When examining the traditional approach, the relevant methodological concerns related to the reliability and sources of intuitions. As handling this objection is explicitly the aim of the accounts we have been considering, this can be left aside for the moment. For the experimental approach, I identified three methodological restrictions. First, the rejection of moral expertise cannot be justified. Second, ethics is properly understood as a prescriptive endeavor, not a descriptive one. Finally, substantive moral theories are not subject to empirical debunking through empirical study of folk intuitions. The last restriction comes in the form of the fact/value distinction, defended in the previous chapter. However, as this is a concern that only applies to moral philosophy, we should not expect to see it addressed here.

As we have already seen, this account is compatible with and defends the notion that philosophers can have expert intuitions, so the first restriction is satisfied. Similarly, our second restriction is also satisfied. Although this account makes use of empirical data and descriptive accounts of people's intuitions, it nevertheless aims at picking out what knowledge actually is and what our epistemic norms should be, and not just at figuring out how people tend to think about these topics. Finally, this approach also does not commit to the idea that substantive philosophical theories can be debunked by empirical examination of folk intuitions. Although Mortensen and Nagel allow for the possibility of a restrictionist use of empirical data, it is important to note what they see as the role of such data. According to them, "[t]he goal of restrictionism is, at the end of the day, not to commit armchair theorizing to flames but to advocate the necessity of empirical work if any evidential role for intuitions is to be salvaged from the fact that our intuitions are problematically sensitive to irrelevant factors" (64). There is a lot to say about this claim, and we'll explore it further in a moment, but for now, note that even this restrictionist

project does not seek to debunk. So, this account covers the methodological restrictions we have previously identified.

This approach also starts to show us a way to balance empirical and traditional approaches. There are a few key methodological lessons to take from this. The first and potentially most important one is this: some questions – including questions we come across when investigating a non-empirical matter – are empirical questions, and answering them requires empirical work. In particular, questions about the nature of our intuitions are helpfully answered through empirical inquiry. At the same time, however, this in itself cannot be the end of philosophical inquiry, but rather is only one step of the process. As Mortensen and Nagel quite correctly put it, “Careful experimental work will allow us to develop a better understanding of the mechanisms driving our judgments. This understanding will, in turn, help us develop more accurate, powerful, comprehensive philosophical theories in combination with extant, armchair modes of theorizing” (66-7). To put it a bit differently, it is useful to think of it in terms of the contrast with the empirical approach. To the empirical approach, the methods of cognitive science are a tool that allows us to get at substantive philosophical answers in the form of empirical data. On this account, however, these methods act as a tool that can help to clarify issues we are facing and, can lead to substantive philosophical answers when combined with our other philosophical tools, including conceptual analysis, intuitions, etc.

The second major lesson we can draw from this concerns the status of intuitions. Like the empirical approach, this account recognizes that epistemic intuitions are prone to error and potentially unreliable. Unlike the empirical approach, however, this account does not treat that as a reason to abandon our use of intuitions. Instead, this issue is taken as an opportunity to dig deeper into the nature and sources of our epistemic intuitions to determine when we should and should not be concerned about potential sources of error and bias, and when they can be treated as reliable. And so, from this perspective, empirical data on possible sources of error in our intuitions serve a therapeutic purpose, as part of a “project of determining where the intuitional signal does, or does not, fall too much prey to noise” (Mortensen and Nagel, 67). The main result of this difference in approach is in how it handles the discovery of some source that interferes with our

intuitions. While the empirical approach may point out that epistemic intuitions are vulnerable to framing effects and conclude that this gives us reason to be skeptical about the use of intuitions to determine philosophical truth, this approach takes that same starting point and tries to determine the scope and extent of this effect, and asks how we might control or compensate for it.

The question that motivated this section was, “Can we take anything from epistemology that can help clarify an appropriate methodological approach for moral philosophy?”, and it seems that the answer is yes. As we have seen, this “armchair-friendly” approach to experimental philosophy is compatible with the methodological concerns and restrictions we developed in the first three chapters. More importantly, the further methodological lessons taken from this epistemological approach are at least *prima facie* useful for laying out further methodological principles for moral philosophy. The biggest challenges for us thus far have been determining what (if any) use can be made of empirical data, and what the proper role of intuitions is. This approach suggests a possible answer to both of these challenges.

## 5.2 Moral Intuitions Revisited

With all of this, we are now – at last – in a position to sketch out a methodological approach to moral philosophy that is compatible with adopting an overall naturalist perspective. We have a set of characteristics that we know this approach must have. In order to be able to actually answer the sorts of questions we’re asking when we ask what makes an act morally right or wrong, we need to be able to offer a prescriptive account of morality. That is to say, if all we have to offer is a description of what most people believe, or how moral beliefs diverge across cultural, gender, or socioeconomic lines, we are failing to answer the questions that motivate moral philosophy. At the same time, we need to be aware that the intuitions that have long been the main currency of moral philosophy are not as reliable as we would like them to be. They are subject to interference from cognitive biases and evolutionary influences that are apparently unconnected to moral truth. Finally, because of the fact/value distinction, for any account of ethics to be effective at being prescriptive, it must be prescriptive throughout. That is

to say, we cannot arrive at a prescriptive account of morality through purely descriptive premises. With all of that in mind, let's start bringing this together.

Our moral intuitions are subject to influence from seemingly-irrelevant factors such as cognitive biases, and an evolutionary history that pushes our moral sentiments towards what is adaptive, rather than any sort of moral truth. This much we know. The question we face is, what should we do with that information? As we have already seen, the response from the empirical approach is that this means we should be skeptical of our intuitions, and should not treat them as a solid basis for philosophical theorizing. As we have also seen, however, intuitions are indispensable to moral philosophy as it has traditionally been done. So, our situation appears to be that moral intuitions are both highly problematic and yet entirely necessary if we want to do ethics well.

The importance of determining the right attitude to have towards moral intuitions cannot be overstated. The empirical approach faces the difficulties it does in no small part because it does not properly address the possibility of expertise in moral intuitions, and regards unreflective folk intuitions as being able to answer substantive philosophical questions. At the same time, the major stumbling point for the traditional approach arises from placing too much faith in the reliability of intuitions, and consequently affording them too much weight in our deliberations. So, figuring out how to properly balance our reliance on intuitions is absolutely vital for any account of moral methodology. There are two routes we can take that balance these concerns: we can reject the skepticism and biofatalism of the empirical approach, or we can adopt a fallibilist approach to intuitions.

### 5.3 Resisting Cognitive Bias

The most basic response to the pessimism of the empirical approach is to show that our intuitions are not as unreliable as the empiricists claim they are. There are several ways that ethicists can accomplish this, depending on exactly what objection is being raised. The first objection I considered was the case of objections raised by citing cognitive biases like the framing effect. In chapter 1, I mentioned Shafer-Landau's criticisms of Sinnott-Armstrong's *master argument*. He objects that arguments of this form do not appropriately distinguish between two propositions: that a small number of our intuitions



are vulnerable to the effect in a large number of circumstances, or that many or most of our intuitions are vulnerable to it (2008, 86). He further argues that the empirical studies supporting the *master argument* and establishing the presence of cognitive biases only reveal “that some such processes are unreliable” (ibid., 88), and that, “I don’t know whose processes, or which processes, are likelier than not to lead to error” (89). At the time, I said that this response was insufficient to assuage concerns about the vulnerability of our intuitions to seemingly irrelevant factors, as that very uncertainty he points to is why we need to treat our intuitions as generally suspicious. We know that our intuitions can be affected by cognitive biases, but we do not know the exact circumstances under which this occurs. So, for any given intuition, we cannot know whether or not it has been led astray by irrelevant influences, so we must be skeptical of all of them. Now, however, we are in a position to answer that challenge.

As we saw from the epistemological example, the mere fact that our intuitions are vulnerable to cognitive biases does not need to be the end of our inquiry, nor a reason to reject the use of intuitions entirely. Instead, we can see it as an opportunity for further investigations into the nature and reliability of our intuitions. As Shafer-Landau points out, there is a big difference between showing that some of our intuitions are vulnerable to these outside influences and showing that all or even many of them are. In the latter case, we need to be careful of any use of intuitions as substantial evidence. In the former case, however, we only need to be cautious when the circumstances under which our intuitions are unreliable obtain.

Again, a comparison to visual perception is useful here. We know that we can be tricked by all sorts of optical illusions, but only the most radical of skeptics thinks this is reason to reject the use of our eyes as a tool for gathering evidence. Instead, we’ve looked deeper into optical illusions to determine when they occur, how they trick us, and so on, allowing us to – for the most part – be able to distinguish between situations in which we should be concerned about potential optical illusions, and ones in which we do not need to. We can do the same for moral intuitions.

Our current situation is akin to having recently discovered that optical illusions exist. We know that external influences can shape our intuitions, and it now falls to us to determine the extent and scope of this vulnerability. In Sinnott-Armstrong's *master argument*, he asserts that we ought to know that moral intuitions are "subject to framing effects in many circumstances" (2008, 52), which is to say that enough evidence exists that demonstrates the existence of framing effects that we are failing to live up to our epistemic responsibilities if we do not know this. Although I would agree with this, I would also suggest that it needs to be pushed a step further. We have an epistemic duty to be aware of the existence of framing effects, but as philosophers – and especially philosophers who work heavily with intuitions – we also have an epistemic duty to investigate and understand the scope and limits of these effects. Unfortunately, this cannot be an a priori task, handled by intuitions. By definition, these effects are not ones of which we are aware, nor do they appear to us when we reflect on our intuitions. But this is where the tools of the empirical approach to ethics are valuable. As we saw in Mortensen and Nagel's analysis, we can narrow in on the set of situations in which framing effects apply – and thus know when our intuitions are relatively insulated by these effects – in the same way the existence of framing effects was discovered in the first place: empirical investigation. As I have argued, the fundamental questions of moral philosophy are not empirical, but questions about exactly when our intuitions are prone to go awry, and thus how we can take steps to control for that, are very much empirical questions. The empirical approach overreaches and overstates just how much of a problem the information we currently have on cognitive biases causes for our use of intuitions, but this does not mean that its tools are entirely without value; in this case, they are of great value if we want to avoid error in our use of intuitions as much as possible.

There is a further upshot of gathering more data about the particular situations in which cognitive biases are especially prevalent. If we find that there are situations in which framing effects are weak or absent, then we know that intuitions formed under

those circumstances are reliable.<sup>97</sup> Perhaps more importantly, it can also allow us to reflect on and be more cautious of our intuitions when we know we're facing a situation in which framing effects are particularly strong. I will discuss reflection on intuitions as a tool to help adjust for potential error further below, but for now, let's return again to the comparison to optical illusions. Knowing what causes optical illusions (odd lighting conditions, strange arrangements of shapes, forced perspective, etc.) allows us to know when what we are perceiving is likely *not* the result of an optical illusion, to be sure, but it can also allow us to recognize when what we see might be illusory, and try to "see through" it. If I know about the Müller-Lyer lines,<sup>98</sup> for instance, then when I see a pair of arrows that resemble the arrows used by this illusion, I know that I should be cautious of my snap judgment that the lines are of different lengths, and look more carefully before forming my final judgment.

## 5.4 Biofatalism

The other major objection to the traditional approach to ethics comes from Street's Darwinian Dilemma. As discussed in the first chapter, Street argues that our basic evaluative tendencies – "an unreflective, non-linguistic, motivational tendency to experience something as 'called for' or 'demanded' in itself, or to experience one thing as 'calling for' or 'counting in favor of' something else" (2008, 119) – are directly influenced by selection pressures, and these basic tendencies have strong influences on our fully-formed judgments (119-20). We then either have to assert that these selection pressures resulted in a set of moral intuitions that have, coincidentally, tracked moral truth – which is implausible and difficult to justify – or we admit that there is no relation between the origins of our moral judgments and any moral truth that may exist, and the influence of natural selection on our moral judgments is "purely distorting" (121). And so, we have an evolutionary debunking argument against ethics.

---

<sup>97</sup> Or, at least, as reliable as any intuition can be.

<sup>98</sup> Parallel lines and arrows, one of which appears to be longer than the other, despite the fact that they are the same length.

From a purely traditional point of view, it is difficult to find a way out of Street's dilemma. The most basic understanding of evolution by natural selection tells us that everything gradually evolved and changed, with adaptive traits that promoted survivability being passed on to future generations, and traits that harm survivability dying out. And so, we might conclude, any trait that has survived the long, slow process of evolution must have been selected for in the first place because it was evolutionarily adaptive. As it happens, however, a more sophisticated understanding of biology does not really support this understanding. This sort of raw adaptationist understanding of evolved traits is now largely understood as being overly simplistic, as it overlooks the possibility of various traits being not directly selected for, but rather a by-product or side-effect of some other trait that was selected for.<sup>99</sup> A purely adaptationist perspective struggles to explain some traits which do not seem to provide any obvious advantages for survival, and may in fact cause problems for organisms that have those traits. It is difficult, for example, to explain how depression and suicide can be understood as promoting survival and reproduction.

This is the problem with Street's account. Ultimately, the assertion that our understanding of evolutionary pressures shows that our moral judgments track what is evolutionarily adaptive rather than any sort of moral truth is not adequately supported by the data. It relies on an oversimplified, overly-adaptationist account of biology and evolution. Absent some evidence that our evaluative judgments are a direct result of natural selection, as opposed to a by-product of something else (our ability to reason and reflect, for instance), the central claim of Street's argument is just one possible hypothesis. If it turns out that our moral sentiments really are directly selected for, and merely push us in directions that make it more likely for our genes to be passed on, then we need to address that, but whether or not this is the case cannot be meaningfully determined from the armchair. This is ultimately an evolutionary biology question. So, once again, the tools of empirical ethics are useful here; to get at an answer to this

---

<sup>99</sup> Cf. Gould and Lewontin (1979)

question, we need more empirical data, and working with the scientists who are able to investigate these sorts of questions is how we can accomplish this.

It is worth noting, in addition to this, that even if Street's hypothesis is correct, and our evaluative tendencies are simply a result of selection pressures, this does not necessarily mean we need to abandon our reliance on our intuitions. Certainly, our evolutionary history has primed us with certain preferences and attitudes, but it is a mistake to think that that is all there is to the story. This sort of biological determinism – or biofatalism – is akin to the physical determinism that has long been at the core of debates in free will, except it points to our biological/evolutionary drives as an inescapable motivator for our behaviours instead of the overall deterministic nature of the universe. Much like these traditional deterministic perspectives, biofatalism is certainly a defensible position, but it is also not undeniably correct and can be coherently rejected.

A full argument against biofatalism is beyond the scope of this project. For a detailed argument against biofatalism, see Barker (2015), who argues that the notion that we are strictly determined to behave in certain ways or have certain social structures as a result of the ways in which evolution had shaped our worldview is simply not supported by the evidence, and humans and human society are quite adaptive to all sorts of change, including ones that seem to go against our “evolutionary nature.” For our purposes, however, this briefer argument will suffice: we have a pretty good understanding of how natural selection has influenced us, and whether or not we act on those basic drives is up to us. We can know, for instance, that natural selection tends to cause us to have an automatic preference for people who look similar to us, but at the same time, we rightly criticize anyone who uses that fact as an excuse to justify their racism. Even if that automatic preference exists, whether or not we go along with it is up to us. On top of this, it is also true that many of us regularly make choices that go against our evolutionary drives. At the most basic level, natural selection is about successfully ensuring that reproduction leads to perpetuating genes.<sup>100</sup> Despite this drive, many people make

---

<sup>100</sup> It doesn't matter how good any given trait is at helping an organism survive if that trait is not passed on.

choices that go directly against it. It is not at all uncommon for affluent people to choose not to have children and to take steps to ensure that they do not do so, and many others choose lives of celibacy. This is not motivated by, for instance, a scarcity of resources, health concerns, or other factors that could negatively affect the viability of potential offspring. Instead, the motivation is often as simple as not wanting to have children, or placing other priorities above reproduction.

Of course, we can tell adaptationist stories about any of these choices. We could explain religious celibacy by explaining that belief in a higher power was useful for bringing people together, promoting group coherence and mutual protection, and thus made it more likely that members of that religious tribe would survive over groups that didn't have similar motivation to stick together. We could then argue that religious celibacy is a result of this, and although it is not necessarily good for the reproductive success of the person who adopts the practice, it is good for the group of which they are a member, and thus is evolutionarily adaptive. Similarly, we could offer an explanation for the decision of affluent individuals to avoid having children in terms of other factors that are evolutionarily adaptive. As psychologist Rod Martin points out, however, “evolutionary theories [. . .] need to provide testable hypotheses making them potentially falsifiable so that they can be more than merely ‘just so’ stories” (Martin 2007, 188). While the adaptationist account offers a legitimate possibility, its proponents cannot just help themselves to the assumption that it is correct. It is up to the adaptationist to provide evidence for their account, and explain how it is more useful or accurate than other accounts, rather than just showing that it is plausible.

## 5.5 Fallibilism and Thought Experiments

Let's suppose, however, that empirical investigation does not support the account I have sketched out here. Suppose we looked further into cognitive biases and framing effects, and discovered that nearly every intuition is thoroughly influenced and corrupted by irrelevant factors in ways we can neither predict nor control for, and further developing our understanding of the influence of natural selection on our intuitions and choices

revealed a deep connection that we were nearly unable to overcome.<sup>101</sup> Would we be left with no route to investigate the sorts of moral questions in which we are interested? In other words, even if moral philosophy is not an empirical endeavour, is pursuing it possible only if the right set of empirical conditions obtain? Answering this question brings us to our final methodological tool: fallibilism about our intuitions.

The simplest possible statement of the concerns raised by the empirical debunking argument is, “What if our intuitions are wrong?” The central concern of both major objections is that there is a good chance that our intuitions are mistaken. But this is only of vital importance to moral philosophy if our approach requires that our intuitions be unassailable. If awareness of the potential for our intuitions to be mistaken or misled is built into our approach to ethics, this problem disappears. The best-known version of such an approach would be Robert Audi’s fallibilistic intuitionism (cf. Audi 1998, 2004, 2015). Audi holds that even if our intuitions are self-evidently justified, that does imply that we are indefeasibly justified in believing them, claiming that, “even the necessary truth of what one believes does not imply that one’s justification for it is indefeasible. We can cease to be justified in believing even a theorem that is both necessary and a priori because our ‘proof’ of it is shown to be defective.” (Audi 1998, 19). This sort of awareness of potential error is exactly what appropriately responding to empirical criticisms demands. This is only one possible approach, however. We could also defend, for instance, a coherentist-style approach, according to which we are concerned with creating a coherent overall account, rather than being especially concerned with whether or not any particular individual part of the picture is necessarily correct. Alternately, we could be foundationalist about our moral system: there are some basic moral facts about which we cannot be wrong, and our moral theory is derived from a combination of those and our other (potentially fallible) intuitions.

Any of these approaches are viable, and there are certainly other approaches to ethics that could be constructed along these lines. I am not advocating any particular

---

<sup>101</sup> This possibility seems extremely remote, but if I want to maintain my claim that these are – at least in part – empirical questions, I cannot rule out that this is a possibility.

approach here. Rather, the concern for our purposes is that whatever approach is taken is one that takes the fallibility of our intuitions seriously. If our account of morality does not require certainty in our intuitive judgments, then it is less important that our intuitions can be disrupted by certain sources of error. This does not entirely remove our epistemic obligation to investigate the sources of error for our intuitions, however. It is still important to know how outside factors influence our judgments, and how we can mitigate that influence to arrive at more accurate judgments. We just don't want our methodology to require that our intuitions be infallible.

With these fallibilist approaches in mind, let us return to thought experiments. Earlier, I laid out four categories of thought experiments, and argued that two of these categories – ones that provide intuitive support for a theory, and ones that provide an intuitive counterexample to a substantive moral theory – are methodologically suspect. What we have seen since then does not change anything about the issues facing moral counterexamples, and we should still be wary of thought experiments that aim to provide one. It does, however, give some reason to re-evaluate thought experiments as a tool for providing intuitive support for a theory. The major objection to that type of thought experiment came from concerns about the reliability of intuitions, a problem to which we now have at least the beginnings of a response. In addition to this, we can also now begin to answer another question that has been raised about the use of thought experiments in moral philosophy.

Moral thought experiments cover a wide range of subjects, from the relatively real-world scenario of a runaway trolley,<sup>102</sup> to truly outlandish and difficult to imagine scenarios, such as people born with two pairs of eyes who expel the second pair from their shoulders at the age of twenty (Lippert-Rasmussen 2008, 98). One question that has been raised about these thought experiments is exactly how outlandish we can make the situations described in our thought experiments before we are no longer able to make use of the intuitions we generate from them. It has been argued that we ought to avoid

---

<sup>102</sup> Admittedly, some of the more elaborate trolley problem-like situations can begin to move beyond the ordinary world.



imaginary cases that are too outlandish, as we cannot properly imagine them, and so our ability to properly apply our moral judgments is compromised.<sup>103</sup> But if we conceive of our attempts to probe our intuitions about outlandish cases from within a fallibilist framework, we can understand this a bit differently.

Whatever else we may think about our moral intuitions, some are more fundamental than others. These more fundamental intuitions can be treated as a sort of moral bedrock. That is, something that is a prototypical case of moral rightness or wrongness, such that if it is not morally right/wrong, then nothing is.<sup>104</sup> Examples like throwing infants onto a bonfire for fun would be among these intuitions: if throwing infants onto a bonfire for fun is not morally wrong, then our moral discourse is so fundamentally misguided that we cannot meaningfully engage in the practice. If we want to expand beyond just these prototype intuitions, and build our understanding of morality from more than just the notion that some things are trivially right and some things are trivially wrong, then we need to also make use of intuitions about which we are less certain. As I have already said, our epistemic position does not afford us certainty about the accuracy of our intuitions, but this does not mean that we lack any other method to test our intuitions and theories.

Unless we are willing to adopt a paraconsistent logic, a minimal requirement of any fully fleshed out ethical theory is that it be consistent; it should not require its adherents to simultaneously hold  $p$  and  $\sim p$ . If we want to test our intuitions for consistency, outlandish cases are particularly helpful. Moral thought experiments can help us by “vary[ing] a given case in precise ways in order to isolate morally relevant factors” (Elster 2011, 242). In general, one of the reasons thought experiments are valuable is that they can help us control for variables that we otherwise couldn’t isolate in real-life situations, as ordinary situations are messy, with a lot of factors all competing for

---

<sup>103</sup> For an example of this argument and overview of several different approaches taken to it, see Elster 2011.

<sup>104</sup> Or, at least, nothing we can meaningfully talk about.

moral relevance. Even the philosopher's refrain of "all other things being equal" rarely (if ever) describes an actual situation. So, to get at a situation in which the only variable is the one we are testing for moral relevance, we need to use artificial cases. Particularly outlandish cases – cases that describe situations which could never occur in the real world – are especially useful, as they can isolate properties that cannot be isolated in realistic scenarios. For instance, we can have a situation in which "the feature X, singled out as morally salient by principle P, might in this world always co-exist with feature Y. However, it might actually be Y which is morally relevant, and we can only see this from outlandish examples where X and Y come apart" (Elster 2011, 244). In such cases, our intuitions may be less clear, and we may be mistaken in our intuitive judgments, but this is a problem we face with our intuitions in general. What we are looking for in these cases isn't certainty, though. We are looking to see if our judgment in such cases is consistent with our other judgments. If they are, then all is well. If they are not, then this presents us an avenue for further inquiry, as we know our judgments must have gone wrong somewhere. Either way, thought experiments using outlandish cases are of use to us.<sup>105</sup>

It is worth taking a moment to note that adopting a fallibilist attitude towards our intuitions is not incompatible with pushing back against the pessimism towards intuitions expressed by proponents of the empirical approach. We can simultaneously hold that we should be cautious about potential errors in our intuitions while also looking to identify and eliminate (if possible) the sources of those errors. Conversely, even if we manage to identify and correct for various sources of error in our intuitions and reach a point that there are no known systemic sources of error which we have not managed to either eliminate or compensate for, epistemic humility still requires us to bear in mind that it is possible that we are wrong. This is especially true when we have previously dealt with empirically-discovered sources of error. Even if we have dealt with the sources of error

---

<sup>105</sup> It is possible to have a case that is so outlandish and so difficult to imagine that we have no firm intuitions about it whatsoever. In such a case, we cannot take anything of use from it, but these situations are bizarre enough that it should be evident to us in advance that they are not going to provide any useful evidence.

of which we are aware, we must grant the possibility that there are other empirically-discoverable confounds on our intuitions which we have yet to discover. As we have seen from the above discussion, this possibility cannot be ruled out a priori, and so we must allow for it. Therefore, these two responses should be seen not as an either/or dichotomy, but as two complimentary strategies that each give us unique tools for dealing with error and uncertainty.

## 5.6 The Methods of Ethics

The only question that now remains is, what does this all add up to? We have examined various methodological objections to both the traditional and empirical approaches to moral philosophy, and considered what sort of methodological restrictions must be in place for an approach that will avoid the problems we identified. To quickly recap one last time, in order to satisfactorily answer the sorts of questions we are asking that lead to moral inquiry, our account must be prescriptive. That is, it must be able to give an answer to a question of the form, “I know that I/other people think x is right, but is it really?” without just reporting on what people think, or what the sources of certain intuitions are. We also know that, in order to be adequately prescriptive, our account cannot rely on purely descriptive premises then attempt to reach a normative conclusion. As we saw in our discussion of the is/ought gap, attempts to derive normative conclusions from non-normative premises simply do not work, and so we must rely on normative premises in order to create an adequately normative account of morality. Finally, any fully satisfactory account of ethics will have to take potential sources of error for our intuitions seriously. Fortunately, as we have seen, we have tools to help us deal with these issues, and help determine when our intuitions are vulnerable to outside interference and when they are not.

It is tempting to conclude from this that the traditional approach is vindicated. The most significant challenge posed concerned the reliance of the traditional approach on intuitions as the major source of evidence. If, as I have argued, however, we have tools for re-establishing the reliability of intuitions, this challenge seems to have disappeared. And so, we might think the traditional approach is able to address all of our concerns. Unfortunately, this conclusion would be a bit too hasty. Properly addressing the concerns

raised by empirically-minded philosophers requires using more than just the methodological tools of the traditional approach. Our approach to ethics is helped by not just being aware of empirical research into cognitive biases and the evolutionary origins of our moral sentiments, but by actively engaging with these research projects to push our knowledge forward and develop our understanding of where and how our a priori moral theorizing interacts with and is potentially affected by empirical work.

Similarly, we cannot think that the empirical approach alone can lead to an adequate account of morality. As we have seen, the tools of the empirical approach lack any means of providing a prescriptive account of ethics. At most, they can give a thorough account of what people's moral attitudes actually are, but this fails to answer the sorts of questions we are asking when we ask whether or not something is right. The best hope for the empirical approach was the potential to limit our set of options for plausible moral theories by finding empirical defeaters for moral theories, but that project is – at best – merely duplicating work that can be done at least as well by traditional approaches to ethics.<sup>106</sup> The empirical approach does provide us with valuable criticisms of the traditional approach, and were it to be the case that there was no possible way to actually craft a satisfactorily prescriptive account of ethics, this approach would be the only viable alternative. As it stands, however, the empirical approach cannot offer a substantively prescriptive moral theory, and so it cannot be our approach.

Instead, what we need to adopt is a hybrid approach that balances the strengths of both the traditional and empirical approaches. The exact shape this takes is less important, and there are likely many different accounts of moral methodology that will satisfy our desiderata. What is required is that we take seriously not only the challenges posed by empirical data, but also – as we saw in the example from Mortensen and Nagel above – the potential for empirical data to “help us develop more accurate, powerful, comprehensive philosophical theories in combination with extant, armchair modes of theorizing” (Mortensen and Nagel, 66-7). This requires substantive engagement with

---

<sup>106</sup> Cf. Ch. 2, §2.4

empirical research on the part of moral philosophers. The exact form this engagement takes can vary, and can include anything from simply remaining aware of important new developments in cognitive science and evolutionary biology to active collaboration with scientists or (when we are dealing with something properly understood as an empirical question) designing and conducting experiments. This hybrid approach requires not only the analytic skills that have always been required by the traditional approach, but also the scientific knowledge and literacy employed by empirical and experimental philosophers.<sup>107</sup>

Mortensen and Nagel dubbed their approach “armchair-friendly experimental philosophy,” but this term is not quite right for the approach I have sketched here. Such a name implies that the approach is primarily experimental, but does not rule out the possibility of armchair intuitions. My proposed approach, on the other hand, is ultimately more similar to the traditional approach than it is to the experimental one. The major source of evidence for our theories is our moral intuitions and conceptual analysis based on those intuitions. Empirical studies are certainly relevant, but they are not the primary source of evidence as they are in the empirical approach. Instead, the empirical studies are used to support, clarify, or provide insight into the intuitions that drive our moral inquiry. The name, “empirical-friendly armchair philosophy” would be more appropriate, but this is also not quite right. This approach is more than just a slightly altered version of the traditional approach which is better able to respond to empirical data; making use of the data is at the core of the approach. Given this, a name like, “empirically-informed moral philosophy” seems most fitting. Science provides us with data and informs our inquiry, to be sure, but deep moral theorizing is still ultimately accomplished as it has always been: rational reflection on our intuitions.

Given that this empirically-informed approach bears strong overall similarities to the traditional approach, it should come as no surprise that some moral philosophers have

---

<sup>107</sup> This is not to imply that moral philosophers have been scientifically illiterate up to this point. As we will see shortly, this is certainly not the case. Rather, my claim is that the traditional approach has not generally *required* much in the way of scientific literacy, as it largely an a priori project.

advocated for approaches to moral philosophy that are compatible with this methodology. In chapter 1, I briefly mentioned Robert Noggle's nature to morality approach, and we should now look at it in more detail. According to Noggle, his account, "evaluates a moral claim or moral theory on the basis of its relation to some (alleged) facts about the kind of beings we are" (532), motivated by the idea that, "the moral norms governing something must depend at least in part on what that something is" (533). This is to be understood in contrast with more purely intuitionist theories, which are not explicitly tied to any particular theory about human nature. Noggle identifies two methodological assumptions that the plausibility of his approach rests on. First, his account assumes that "a moral theory can be 'linked' to a nature-claim, even if it neither asserts nor denies that claim" (534). The second assumption is that "the plausibility of a moral theory can be affected by whether or not it reflects the truth about human nature and/or the nature of persons" (*ibid.*). The specific proofs he offers aren't of particular importance, since it is the overall methodology is what is of interest. So, we can leave his defenses of these assumptions aside.

Notably, these assumptions match up perfectly with the methodology that has been laid out here. The second assumption encapsulates exactly the objection from Street that we are trying to address: accounts of morality that do not accurately reflect how our evolutionary history has shaped our intuitions and preferences are less plausible than ones that do get this evolutionary story right. Similarly, the notion that a moral theory can be "linked" to a claim about nature is reflected in how we have attempted to deal with errors stemming from cognitive bias. Sinnott-Armstrong's master argument establishes that we have an epistemic responsibility to know that framing effects exist and be aware of their impact on our intuitions. Practically speaking, if we are aware of the presence of framing effects but our moral theorizing does not reflect the fact that these can cause our intuitions to go awry, this amounts to denying that the impact of these effects is relevant. Noggle's main idea is that a proper understanding of ethics needs to be connected to our understanding of human nature. This is also important to the empirically-informed approach. If we want to give a fully satisfactory account of ethics, we cannot be entirely divorced from psychological and evolutionary theories and how they bear on our understanding of morality. Noggle does not explicitly tie the theories of human nature

that he cites in his approach to purely scientific theories of human nature (he also allows for the possibility of a priori, philosophical accounts), but such scientific accounts are certainly compatible with his approach.

Noggle is far from the only moral philosopher whose approach to ethics is compatible with the empirically-informed approach. Allan Gibbard's expressivism (Gibbard 1992; 2003) is deeply informed by work in psychology, biology and evolutionary game theory, while remaining a prescriptive theory. Recently, Peter Railton (2014) has also examined empirical approaches and objections to traditional moral philosophy, and looks at how empirical data may help us to deal with these objections. Importantly, though, Railton's conclusion – like that of Mortensen and Nagel, or the empirically-informed approach – is that this empirical research can be used to improve our use of intuitions, not to debunk. Specifically, he claims that research into the sources and nature of our intuitions can give us, “a firmer theoretical and evidential basis upon which to explore the nature and informativeness of the ‘moral intuitions’ upon which we rely extensively and inevitably” (Railton 2014, 859). In other words, we cannot abandon our use of moral intuitions; we use them, “extensively and inevitably.” What we can do is use the data to give us “a better-grounded epistemology of moral intuition, enabling us to assess more critically and confidently intuition's diverse roles in moral thought and action” (ibid.). So, again, the methodology employed here is to use intuitions to drive our philosophizing, but attend to and incorporate empirical data to help improve our confidence in our intuitions, which is exactly what the empirically-informed approach advocates.

Considering the common ground the empirically-informed method shares with the approaches mentioned above, we may now ask how this differs from the general methodology of reflective equilibrium. In the first chapter, I said that the empirical and traditional approaches could be understood as disagreeing less about the method of reflective equilibrium and more about what weight should be given to various pieces of evidence. The empirical approach weighs scientific data highly and gives very little weight to intuitions, while the traditional approach privileges intuitions and gives less weight to empirical data. Given this starting point, it would not be surprising if my

proposed empirically-informed approach did not differ greatly from the general methodology of reflective equilibrium. The empirically-informed approach is certainly compatible with existing approaches to reflective equilibrium. It focuses more on current information in biology and cognitive science and requires moral philosophers to be more directly engaged with these sciences than traditional approaches do, and thus requires that they be given more weight in our deliberations, but does not require radically rethinking the method of reflective equilibrium. In this sense, the empirically-informed approach is perfectly compatible with reflective equilibrium, mostly affecting how much weight we give different pieces of evidence. That said, it is not the case that the empirically-informed approach should be understood as simply a form of reflective equilibrium. As I mentioned in the first chapter, reflective equilibrium is a coherentist methodology, and while the empirically-informed approach is compatible with a coherentist methodology, it is equally compatible with non-coherentist approaches. The empirically-informed approach need not commit to either a coherentist or non-coherentist picture of moral reasoning. Any approach that takes seriously both our moral intuitions and relevant empirical data in the ways I have described above will be compatible with the empirically-informed approach.

## 5.7 Conclusion

It is an interesting time to be doing work in ethics. Research in cognitive science and evolutionary biology offers us unparalleled insight into the sources of our moral intuitions. While this can present challenges to the ways in which we have traditionally understood the practice of moral philosophy, it also provides opportunities to better understand the factors that shape our intuitions. Neither the traditional approach to ethics nor the empirical approach are fully able to offer a satisfactory methodology for investigating moral philosophy under these circumstances. The empirical approach is unable to offer the sort of prescriptivity that an account of ethics needs in order to answer the sorts of questions that lead us to doing ethics in the first place. On the other hand, the traditional approach, by neglecting to adequately respond to potential sources of error, winds up relying on intuitions that lack the sort of reliability we want them to have in order to serve as the basis of our moral theories.



To resolve this, we need an approach that is able to balance the strengths of both the traditional and empirical approaches without committing itself to the methodological mistakes that make these accounts problematic. Specifically, the account must offer more than just a description of what sorts of moral preferences people actually hold, or how we come to form moral judgments. Rather, it must be able to provide some answer to the question, “What *should* our moral intuitions be in this case?” In order to do this, the argument that establishes our moral theory must itself derive from normative premises, and as such, the purely descriptive data of the empirical approach cannot suffice. Further to this, a successful moral theory must also engage with the sources of error we have identified, working with data from cognitive science and evolutionary biology to better understand how these factors influence our judgments, and how to control for them. In essence, what we are looking for is an approach that balances the use of empirical data against the traditional intuitionist method of moral argumentation. As I have argued, this is what empirically-informed moral philosophy offers. This methodology uses the analytic tools of the traditional approach, but also demands engagement and collaboration with our empirically-minded colleagues in order to give our understanding of the intuitions that are fundamental to our moral thinking a firmer foundation.

Ultimately, this approach is not entirely new. As we saw in the cases of Gibbard, Noggle, and Railton, there are already those who we would consider more or less traditional ethicists who are already engaged with this practice. Work in evolutionary game theory has similarly attempted to give normative accounts while simultaneously expanding our understanding of the potential evolutionary origins of prosocial or moral behaviour. For most traditional moral philosophers, adopting the empirically-informed approach requires relatively little change. What it requires is that ethicists be aware of significant work in cognitive science and evolutionary biology on cognitive biases and moral intuitions. It does, however, also leave open the possibility for ethicists to be much more involved with this empirical side of the project, actively taking part in empirical investigation in ways we would up to now identify more closely with experimental philosophers. With this approach in mind, it is clear that a naturalist worldview does not in fact pose any significant threat to the prospects of continuing to do moral philosophy

and engage with the sorts of substantial moral questions we always have, so long as we are cautious about our methods.

## Bibliography

- Anscombe, G.E.M. (1958). Modern Moral Philosophy. *Philosophy*, 33(124), (1-19).
- Arras, J.D. (2007). The Way We Reason Now: Reflective Equilibrium in Bioethics. In B. Steinbock *The Oxford Handbook of Bioethics* (pp. 46-71). New York, NY. Oxford University Press.
- Audi, R. (1998). Moderate Intuitionism and the Epistemology of Moral Judgment. *Ethical Theory and Moral Practice*, 1, (15-44).
- Audi, R. (2005). *The Good in the Right*. Princeton, NJ. Princeton University Press.
- Audi, R. (2015). Intuition and Its Place in Ethics. *Journal of the American Philosophical Association*, 1(1), (57-77).
- Bar-Hillel, M. (1980). The Base-Rate Fallacy in Probability Judgments. *Acta Psychologica*, 44, (211-233).
- Barker, G. (2015). *Beyond Biofatalism*. New York, NY: Columbia University Press.
- Bertrand, M., and Mullainathan, S. (2004). Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *The American Economic Review*, 94(4), 991-1013.
- Bourget, D. and Chalmers, D.J. (2014). What do philosophers believe? *Philosophical Studies*, 170(3), (465-500).
- Boyd, K., and Nagel, J. (2014). The Reliability of Epistemic Intuitions. In Edouard Machery (Ed.) *Current Controversies in Experimental Philosophy* (pp. 109-127). New York, NY: Routledge.
- Brown, É. (2016). Aristotelean virtue ethics and the normativity challenge. *Dialogue*, 55(1), (131-150)
- Brown, J.R. (2004). Peeking into Plato's Heaven. *Philosophy of Science*, 71(5), (1126-1138).
- Buckwalter, W. and Turri, J. (2015). Inability and Obligation in Moral Judgment. *PLoS One*, 10(8).
- Campbell, R., and Kumar, V. (2013). Pragmatic Naturalism and Moral Objectivity. *Analysis*, 73(3), (446-455).

- Chituc, V., Henne, P., Sinnott-Armstrong, W. and De Brigard, P. (2016). Blame, not ability, impacts moral “ought” judgments for impossible actions: Toward an empirical refutation of “ought” implies “can”. *Cognition*, 150, (20-25).
- Churchland, P. (2011). *Braintrust: What Neuroscience Tells Us About Morality*. Princeton, NJ: Princeton University Press.
- Clipsham, P. (2014). Does Empirical Moral Psychology Rest on a Mistake? *Philosophical Studies*, 170(2), (215-233).
- Clipsham, P. (forthcoming). Is there a Distinct Metaphilosophical Companions in Guilt Argument (and Does it Work)? *Philosophia*. (1-16).
- Cowie, C. (2014). Why Companions in Guilt Arguments Won't Work. *The Philosophical Quarterly*, 64(256), (407–422).
- Cowie, C. (2016). Good News for Moral Error Theorists: A Master Argument against Companions in Guilt Strategies. *Australasian Journal of Philosophy*, 94(1), (115–130).
- Cuneo, T. (2007). *The Normative Web: An Argument for Moral Realism*. Oxford: Oxford University Press.
- Dancy, J. (1985). The Role of Imaginary Cases in Ethics. *Pacific Philosophical Quarterly*, 66, (141-153).
- Dennett, D. (1980). The Milk of Human Intentionality. *Behavioural and Brain Sciences*, 3(3), (428-430).
- Deutsch, L. (2015). Is a Hot Dog a Sandwich? Council Rules Once and for All. *USA Today*, 7 Nov. 2015, [www.usatoday.com/story/news/nation-now/2015/11/07/a-hot-dog-is-not-a-sandwich/75362898/](http://www.usatoday.com/story/news/nation-now/2015/11/07/a-hot-dog-is-not-a-sandwich/75362898/).
- Driver, J (2013). Moral expertise: Judgment, practice, and analysis. *Social Philosophy and Policy*, 30(1/2), (280-296).
- Elster, J. (2011). How Outlandish Can Imaginary Cases Be? *Journal of Applied Philosophy*, 28(3), (241-258).
- Flanagan, O. J., Sarkissian, H. and Wong, D. (2008). Naturalizing Ethics. In W. Sinnott-Armstrong (Ed.), *Moral Psychology, Volume 1: The Evolution of Morality* (pp. 1-25). Cambridge, MA: MIT Press.

- Garber, M. A Hot Dog Is Not a Sandwich. *The Atlantic*, 5 Nov. 2015, <https://www.theatlantic.com/entertainment/archive/2015/11/its-not-a-sandwich/414352/>
- Geach, P.T. (1976). Murder and Sodomy. *Philosophy*, 51(197), (346-348).
- Geach, P.T. (1977). Again the Logic of 'Ought'. *Philosophy*, 52(202), (473-476).
- Geach, P.T. (1982). Moral Autonomy Still Refuted. *Philosophy*, 57(219), (127-9)
- Gettier, E. (1963). Is Justified True Belief Knowledge? *Analysis*, 23(6), (121-123).
- Gibbard, A. (1992). *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Oxford: Oxford University Press.
- Gibbard, A. (2003). *Thinking How to Live*. Cambridge, MA: Harvard University Press.
- Gilligan, C. (1982). *In a Different Voice: Psychological Theory and Women's Development*. Cambridge: Harvard University Press.
- Gould, S.J., and Lewontin, R.C., (1979). The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 205(1161), (581-598).
- Greene, J. D. (2008). The Secret Joke of Kant's Soul. In W. Sinnott-Armstrong (Ed.), *Moral Psychology, Volume 3: The Neuroscience of Morality* (pp. 56-79). Cambridge, MA. MIT Press.
- Haidt, J. (2001). The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment" *Psychological Review*, 108(4), (814-834).
- Haraway, D. (1988). Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies*, 14(3), (575-599).
- Haraway, D. (1989). *Primate Visions: Gender, Race, and Nature in the World of Modern Science*. New York: Routledge.
- Harding, S. (1986). *The Science Question in Feminism*. Ithaca: Cornell University Press
- Harding, S. (1991). *Whose Science? Whose Knowledge?: Thinking from Women's Lives*. Ithaca: Cornell University Press
- Harding, S. (1993). Rethinking Standpoint Epistemology: What is "Strong Objectivity"? In L. Alcoff & E. Potter (Eds.), *Feminist Epistemologies* (49-82). New York: Routledge.
- Harding, S. (1995). "Strong Objectivity": A Response to the New Objectivity Question. *Synthese*, 104(3), (331-349).

- Henrich, J., Heine, S.J., Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33, 61-135.
- Hill, S. (2008). 'Is'-'Ought' Derivations and Ethical Taxonomies. *Philosophia*, 36, (545-566).
- Hill, S. (2009). Good News for the Logical Autonomy of Ethics. *Argumentation*, 23(2), (277-283).
- Hume, D. (1738). *A Treatise of Human Nature*.
- Husi, S. (2013). Why Reasons Skepticism is not Self-Defeating. *European Journal of Philosophy*, 21(3), (424-449).
- Jones, E.E. and Harris, V.A. (1967). The Attribution of Attitudes. *Journal of Experimental Social Psychology*, 3(1) (1-24).
- Kahneman, D., Slovic, P., and Tversky, A. (1982) *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge, UK; Cambridge University Press.
- Kang, S.K., DeCelles, K.A., Tilcsik, A., and Jun, S. (2016). Whitened Résumés: Race and Self-Presentation in the Labor Market. *Administrative Science Quarterly*, 61(3), (469-502).
- Kant, I. (1993). *Grounding for the Metaphysics of Morals/ On Supposed Right to Lie Because of Philanthropic Concerns*. Trans. James W. Ellington. Indianapolis: Hackett. (Originally published in 1785/1799).
- Karmo, T. (1988). Some Valid (but no Sound) Arguments Trivially Span the 'Is'-'Ought' Gap. *Mind*, 97(386), (252-257).
- Kitcher, P. (2006). Four Ways of "Biologizing" Ethics. In Elliott Sober (Ed.), *Conceptual Issues in Evolutionary Biology*, 3rd Edition (pp. 575-587). Cambridge, MA: MIT Press. (Originally published in 1994).
- Knobe, J., and Nichols, S. (2008). An Experimental Philosophy Manifesto. In J. Knobe and S. Nichols (Eds.), *Experimental Philosophy* (pp. 3-14). New York, NY. Oxford University Press.
- Korsgaard, C.M. (1996). *The Sources of Normativity*. Cambridge, MA: Cambridge University Press.
- Lippert-Rasmussen, K. (2008). Against Self-Ownership: There Are No Fact-Insensitive Ownership Rights over One's Body. *Philosophy and Public Affairs*, 36(1), (86-118).
- Longino, H. (1990). *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton: Princeton University Press.

- Longino, H. (1993). Subjects, Power, and Knowledge: Description and Prescription in Feminist Philosophies of Science. In L. Alcoff & E. Potter (Eds.), *Feminist Epistemologies* (101-120). New York: Routledge.
- Mackie, J. (1977). *Ethics: Inventing Right and Wrong*. London: Penguin, UK.
- Maitzen, S. (1998). Closing the 'Is'-'Ought' Gap. *Canadian Journal of Philosophy*, 28(3), (349-365).
- Martin, R.A. (2007). *The Psychology of Humor: An Integrative Approach*. Burlington, MA: Elsevier Academic Press.
- Mortensen, K. and Nagel, J. (2016). Armchair-Friendly Experimental Philosophy. In Justin Sytsma and Wesley Buckwalter (Eds.) *A Companion to Experimental Philosophy* (pp. 53-70). West Sussex, UK. Wiley-Blackwell.
- Nagel, J. (2012). Intuitions and Experiments: A Defense of the Case Method in Epistemology. *Philosophy and Phenomenological Research*, 85(3), (495-527).
- Nagel, J. (2013). Defending the Evidential Value of Epistemic Intuitions: A Reply to Stich. *Philosophy and Phenomenological Research*, 82(1), (179-199).
- Nagel, T. (2012). *Mind and Cosmos*. New York, NY. Oxford University Press.
- Nelson, M.T. (2007). More Bad News for the Logical Autonomy of Ethics. *Canadian Journal of Philosophy*, 37(2), (203-216).
- Nichols, S. (2002). How psychopaths threaten moral rationalist: Is it irrational to be amoral? *The Monist*, 85(2), (285-303).
- Nichols, S. (2008). Moral Rationalism and Empirical Immunity. In W. Sinnott-Armstrong (Ed.), *Moral Psychology Volume 3: The Neuroscience of Morality* (pp. 395-407). Cambridge, MA: MIT Press.
- Noggle, R. (2001). From the Nature of Persons to the Structure of Morality. *Canadian Journal of Philosophy*, 31(4), (531-566).
- Nozick, R. (1974). *Anarchy, State, and Utopia*. New York, NY. Basic Books.
- Okruhlik, K. (1994). Gender and the Biological Sciences. *Canadian Journal of Philosophy*, 24(1), (21-42).
- Olson, J. (2011). In Defense of Moral Error Theory. In Michael Brady (Ed.), *New Waves in Metaethics*. (pp. 62-84). New York, NY: Palgrave Macmillan.
- Pew Research Center (2014). "Belief in God – Religion in America: U.S Religious Data, Demographics, and Statistics." Pew Research Center, <http://www.pewforum.org/religious-landscape-study/belief-in-god/>

- Petrinovich, L., and O'Neill, P., (1996). Influence of Wording and Framing Effects on Moral Intuitions. *Ethology and Sociobiology*, 17, (145-171).
- Prinz, J. J. (2007a). *The Emotional Construction of Morals*. New York: Oxford University Press.
- Prinz, J. J. (2007b). Can moral obligations be empirically discovered? *Midwest Studies in Philosophy*, 31, (271-291).
- Prinz, J. J. (2008). Is Morality Innate? In W. Sinnott-Armstrong, *Moral Psychology, Volume 1: The Evolution of Morality* (pp. 367-406). Cambridge, MA: MIT Press.
- Prinz, J.J. (2009). The normativity challenge: Cultural psychology provides the real threat to virtue ethics. *The Journal of Ethics*, 14(2-3), (117-144).
- Prior, A.N. (1960). The Autonomy of Ethics. *Australian Journal of Philosophy*, 38(3), (199-206).
- Putnam, H. (2002). *The Collapse of the Fact/Value Dichotomy and Other Essays*. Cambridge: Harvard University Press.
- Railton, P. (2014). The Affective Dog and Its Rational Tale: Intuition and Attunement. *Ethics*, 124(4), (813-859).
- Rawls, J. (1971). *A Theory of Justice*. Cambridge, MA. Harvard University Press.
- Rose, D. and Schaffer, J. (2017). Folk Mereology is Teleological. *Noûs*, 51(2), (238-270).
- Rowland, R. (2016). Rescuing Companions in Guilt Arguments. *The Philosophical Quarterly*, 66(262), (161–171).
- Scanlon, T. M. (1998). *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.
- Schwitzgebel, E., and Cushman, F. (2012). Expertise in Moral Reasoning? Order Effects on Moral Judgment in Professional Philosophers and Non-Philosophers. *Mind and Language*, 27(2), (135-153).
- Schwitzgebel, E., and Cushman, F. (2015). Philosophers' biased judgments persist despite training, expertise and reflection. *Cognition*, 141, (127-137).
- Schwitzgebel, E. and Rust, J. (2014). The moral behavior of ethics professors: Relationships among self-reported behavior, expressed normative attitude, and directly observed behavior. *Philosophical Psychology*, 27(3), (293-327).
- Searle, J. (1964). How to derive "ought" from "is". *Philosophical Review*, 73(1), (43-58).



- Shafer-Landau, R. (2008). Defending Ethical Intuitionism. In W. Sinnott-Armstrong (Ed.) *Moral Psychology, Volume 2: The Cognitive Science of Morality* (pp. 83-95). Cambridge, MA: MIT Press.
- Sharot, T. (2011). The Optimism Bias. *Current Biology*, 21(23), (R941-R945).
- Sidgwick, H. (1981). *The Methods of Ethics*, 7<sup>th</sup> ed. Indianapolis, Hackett. (Originally published in 1874, 7<sup>th</sup> ed. Originally published in 1907).
- Singer, P. (1972a). Famine, Affluence, and Morality. *Philosophy and Public Affairs*, 1(3), (229-243).
- Singer, P. (1972b). Moral experts. *Analysis*, 32(4), 115-117).
- Sinnott-Armstrong, W. (2006). *Moral Skepticisms*. New York, NY. Oxford University Press.
- Sinnott-Armstrong, W. (2008). Framing Moral Intuitions. In W. Sinnott-Armstrong (Ed.) *Moral Psychology, Volume 2: The Cognitive Science of Morality* (pp. 48-76). Cambridge, MA: MIT Press.
- Street, S. (2006). A Darwinian Dilemma for Realist Theories of Value. *Philosophical Studies*, 127(1), (109-166).
- Street, S. (2013). Evolution and the Normativity of Epistemic Reasons. *Canadian Journal of Philosophy*. 35. (213-248).
- Thomson, James, and Thomson, Judith. (1964). How not to Derive “Ought” from “Is”. *Philosophical Review*, 73(4), (512-516).
- Thomson, J.J. (1971). A Defense of Abortion. *Philosophy and Public Affairs*, 1(1), (47-66)
- Thomson, J.J. (1976). Killing, Letting Die, and the Trolley Problem. *The Monist*, 59(2), (204-217).
- Tobia, K., Buckwalter, W., and Stich, S. (2013). Moral Intuitions: Are Philosophers Experts? *Philosophical Psychology*, 25(5), (629-638).
- Tversky, A., and Kahneman, D. (1973). Availability: A Heuristic for Judging Frequency and Probability. *Cognitive Psychology*, 5(2), (207-232).
- Tversky, A. and Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157), (1124-1131).
- Tversky, A., and Kahneman, D. (1981). The Framing of Decisions and the Psychology of Choice. *Science*, 211(4481), (453-458).

Wason, P.C. (1960). On the failure to eliminate hypotheses in a conceptual task. *The Quarterly Journal of Experimental Philosophy*, 12(3), (129-140).

## Curriculum Vitae

<b>Name:</b>	Christopher Shirreff
<b>Post-secondary Education and Degrees:</b>	<p>University of Toronto at Mississauga Mississauga, Ontario, Canada 2004-2008 B.A. Honours, Philosophy and English</p> <p>University of Toronto Toronto, Ontario, Canada 2008-2009 M.A., Philosophy</p> <p>The University of Western Ontario London, Ontario, Canada 2010-2019 Ph.D., Philosophy</p>
<b>Honours and Awards:</b>	<p>Ontario Graduate Scholarship 2009 (Declined)</p> <p>Graduate Entrance Scholarship The University of Western Ontario 2010</p>
<b>Related Work Experience</b>	<p>Teaching Assistant University of Toronto 2008-2009</p> <p>Teaching Assistant The University of Western Ontario 2010-2012</p> <p>Instructor The University of Western Ontario 2012</p> <p>Instructor Fanshawe College 2018-2019</p>